Q

Quality Measures

▶ Relative Validity Criteria for Community Mining Algorithms

Quality of Social Network Data

Germà Coenders[1], Tina Kogovšek[2],
Valentina Hlebec[3], and Lluís Coromina[1]
[1]Department of Economics, University of Girona, Girona, Spain
[2]Faculty of Arts and Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia
[3]Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

Synonyms

Accuracy; Reliability; Validity

Glossary

Accuracy Extent to which the perceived (cognitive) network corresponds to the "actual" (behavioral) network

Attenuation Extent to which relationships among variables are lower than they should be as a result of measurement error

Cognitive Interview In-depth interview conducted with the aim of understanding the mental process through which the respondent understands the question and produces the answer

Measurement Assignment of numbers to objects according to rules

Measurement Error Discrepancy between a true and observed characteristic. It may be random (lack of reliability) or systematic (lack of validity)

Operationalization Assignment of empirical indicators (e.g., survey questions) to a theoretical variable to be measured

Pearson Correlation Coefficient Statistical measure of association between two numeric variables which are assumed to be linearly related, ranging from 1 (perfect linear positive relationship) to −1 (perfect linear negative relationship) through 0 (no relationship). Also known as linear correlation or product moment correlation

Reliability Extent to which a measurement yields similar results upon repeated trials

Split-Ballot Experiment Experiment in which the sample is randomly divided into two or more parts, each getting a different form of the questionnaire, a different data collection procedure, etc.

R. Alhajj, J. Rokne (eds.), *Encyclopedia of Social Network Analysis and Mining*,
DOI 10.1007/978-1-4614-6170-8,

Validity Extent to which a measurement converges to what the researcher intends to measure

Definition

Data quality in social network analysis can be understood as the agreement between the measured networks and the true networks. Data quality is an important issue when social network data are collected by means of questionnaires.

In the design of surveys and questionnaires to measure social networks, choices must be made with regard to the wording of questions, the response scales, question order, and the technique for data collection, among others. Each of these choices and combinations of choices leads to different measurement errors, in other words, to a different quality of the obtained data. Data quality is a multifaceted concept including both general terms related to questionnaire data quality (reliability and validity) and specific concepts related to the social network case (accuracy).

Data quality in social networks may apply to:

- Social network characteristics such as network composition (e.g., percentages of friends, coworkers, and family members in one's personal network), density, centralization, or number of factions (high-density subgroups), as in Kogovšek and Hlebec (2008)
- Network member characteristics such as personal background, degree centrality, closeness centrality, or betweenness centrality, as in Zemljič and Hlebec (2005)
- Characteristics of ties between two network members such as frequency of contact, closeness, or geographical distance, as in Kogovšek et al. (2002)

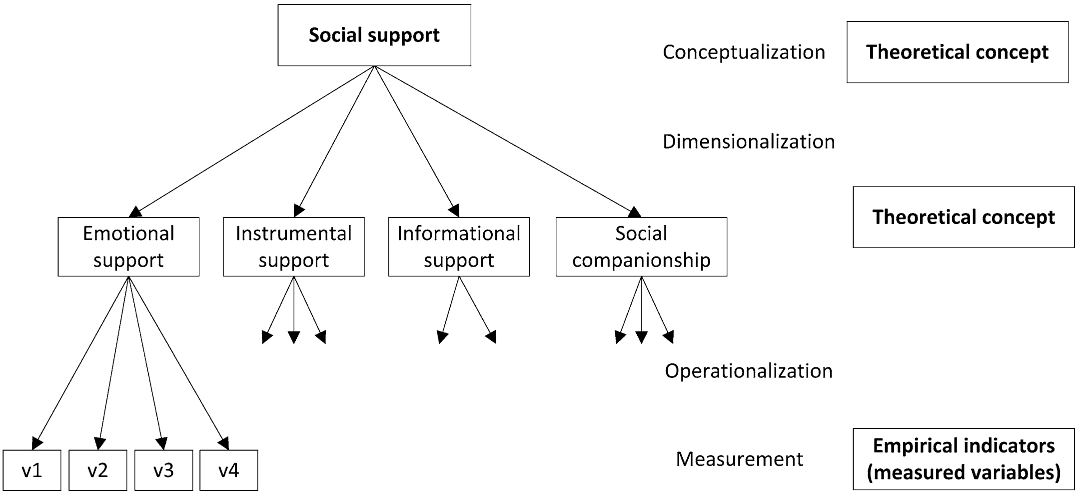
Introduction

Measurement

Measurement is the assignment of numbers to objects according to rules. More broadly measurement could be described as a process of connecting theoretical concepts to empirical indicators – measured variables. Theoretical concepts are usually relatively open, general, and abstract and usually cannot be observed nor measured directly. They can be made measurable (in other words, operationalized) in many different ways, one of which are survey questions. These questions serving as indicators may then be better or worse approximations of the theoretical concepts they intend to measure. Compared to theoretical concepts, empirical indicators are specific and narrow and usually do not represent theoretical concepts in their entirety. In each measurement process, the group of chosen empirical indicators is only a subgroup of all possible indicators of the theoretical concept.

In a somewhat more detailed manner, measurement could be described as a three-stage deductive process (see Fig. 1). Firstly, theoretical concepts and relationships among them are defined. In this stage, different theoretical definitions of the concept are considered, and the most suited to the research purpose is chosen (or, alternatively, if none of the existing definitions is suitable, the researcher formulates his/her own definition). Concepts in the social sciences are usually complex and must then be *dimensionalized*. We could call these different dimensions of a concept theoretical variables. This is still the first stage of measurement which we could call *conceptualization*. In the next stage called *operationalization*, one or preferably more empirical indicators (e.g., survey questions) are sought for each theoretical variable. In the final stage, the actual process of *measurement* takes place (e.g., collecting data from respondents in a survey).

An example of the process of measurement is shown in Fig. 1. The example is taken from the field of social support, one of the most common applications of social network analysis. There are many definitions of social support. Let us here mention only one of the more referential definitions by Vaux (1988) which states that social support is a complex concept with three basic elements: firstly sources of social support (networks



v1. Who are the people you turn to discuss personal problems?
v2. Who helps you bring back your self-esteem when you feel uncertain about something?
v3. Who gives you a feeling that you are respected?
v4. Who do you talk to about important personal matters?

Quality of Social Network Data, Fig. 1 The process of measurement

of people to which we turn to for help in different life situations), secondly types of social support, and thirdly the individual's evaluation of both sources and types (e.g., their availability, adequacy, and sufficiency). So, social support in our example is the theoretical concept in question in the broadest sense of the word. There has been a relative consensus in the research community that there are four basic types (or dimensions) of social support: instrumental (e.g., borrowing money, taking care of children, household help), emotional (e.g., help in different situations such as family or work problems, death of a beloved person), informational (e.g., advice in situations such as looking for a new job or looking for a new apartment), and social companionship (e.g., visiting, going to the cinema). In our example these dimensions of social support are the theoretical variables within the broader concept of social support. In the next stage, these theoretical variables are operationalized in survey questions, such as "Suppose you have a flu, to whom would you turn to bring you some goods from the shop?" or "Suppose you had a quarrel with your husband/wife/partner, whom you would turn to for help?."

Key Points

Accuracy

A concept of quality that is somewhat specific for social network analysis, especially for whole networks, is the concept of accuracy. Accuracy entails the difference between the perceived (also called cognitive) network and "actual" (also called behavioral) network; therefore, the difference between how the network is cognitively represented in the individual's mind and what the network is really like, the actual interactions (behavior) among the network members.

An actual network can be designated in at least two ways:

- By observation
- By aggregation of estimates of the "actual" network (dyadic consensus, member consensus)

According to Pattison (1994) an individual's social environment may affect his/her cognition in at least three ways:

- The environment partly affects which information the individual is exposed to. For instance, if cultural knowledge is being transferred in communication among individuals, the ones in the more central positions in the network (e.g., political leaders, managers) should have more information (knowledge).

- Environment is tied to characteristics and laws of social contacts on which expectations of future events are based therefore causing cognitive bias. For example, comparison of survey data on contacts with direct records of contacts (e.g., automatic electronic records of e-mail exchanges) shows bias in survey data that corresponds to the characteristics of long-term contacts. Freeman et al. (1987) found out that participants in regular meetings remember single occurrences of individual persons poorly but remember regular patterns of presence (who usually attends/does not attend meetings) relatively well. An individual's accuracy of reporting is positively correlated with both frequency of contact and with the individual's general reporting accuracy.
- Cognitive processes may directly include an individual's perception of local social environment. An example could be the so-called theories of consistency – if person A likes person B and there is an important entity C (a thing or a person), there would be a tendency that person A would align his/her perception of C with the supposed perception of B towards C (e.g., "a friend of my friend is my friend").

Cognitive accuracy can also be studied, in other words, to what extent an individual's perception of a whole network agrees with another individual's perception of the same network. According to Krackhardt (1987), individuals with a more accurate perception of the advice network are perceived as having more power. Managers with a more accurate perception of the structure of the network tend to be evaluated more positively by others.

It can also be studied to what extent an individual's perception of a whole network is close to the actual structure. In this case we distinguish between the individual level and dyadic accuracy. The first was studied by Krackhardt (1987). A locally aggregated structure (LAS) means that the relation between two persons is considered to exist if it is confirmed by both. A consensus structure (CS) means that a relation is considered as actually existent if it is confirmed by the majority of network members.

Dyadic accuracy was studied by Bondonio (1998). In comparison with Krackhardt who considers only one estimate of accuracy, Bondonio defines as many estimates of accuracy of an individual respondent as there are network members. An individual respondent can estimate ties between different pairs (dyads) of network members with different levels accuracy which can depend on different factors (e.g., degree centrality of both persons in a dyad, their age, professional position, and geodesic distance [the shortest path between two persons in a network] between them).

Casciaro (1998) focused on LAS of friendship and advice networks. Accuracy of observation of the advice network was positively correlated with the accuracy of observation of the friendship network, with achievement motivation and with indegree centrality in the friendship network and negatively correlated with the hierarchical position for both types of networks. Accuracy of observation of the friendship network was positively correlated with affiliation motivation and with indegree centrality in the friendship network.

Reliability and Validity

A measurement instrument (i.e., a questionnaire) is valid if it really measures the concepts (in this case, the characteristics of network ties, network actors, and networks as a whole) that it is supposed to measure. Reliability is then defined as the ability of the measurement instrument to produce the same results in repeated measurements. Validity concerns the absence of systematic errors which cause a deterministic departure between what the questionnaire intends to measure and what it actually measures. Reliability concerns precision, in other words, lack of random errors and equals the percentage of stable variance.

The lack of reliability results in the attenuation of relationships between variables. For instance, it can be shown that the Pearson correlation between two variables containing random error converges to the true Pearson correlation times the geometric mean of the reliability of both variables. This phenomenon is known as correlation attenuation. The effects of lack of validity are much harder to predict and may both inflate and deflate correlations.

If we relate the measurement process in Fig. 1 to the concepts of reliability and validity, we find that reliability is assessed in the measurement stage and validity pertains to the relationship between conceptualization (what we want to measure) through operationalization (how the theoretical concepts are translated into empirical indicators) to actual measurement (what we actually measure).

Many researchers have been interested in the factors affecting reliability and validity of network measurements.

Kogovšek et al. (2002) studied reliability and validity of tie characteristics associated to data collection mode (telephone versus face-to-face personal interviews) and question order (by questions or by alters). The cheaper telephone mode turned out to be more valid than the personal mode and about equally reliable. Question order by alters (asking all questions for an alter together, before moving to the next alter) produced more reliable tie characteristics than order by questions (asking each question for all alters together, before moving to the next question) and about equally valid.

Related findings regarding question order and data collection mode are that order by questions seems to be working better than by alters in the particular case of a web survey setting (Lozar Manfreda et al. 2004a; Coromina and Coenders 2006). In web questionnaires, graphical design also has quite a large effect on data quality (Lozar Manfreda et al. 2004b; Coromina and Coenders 2006). Reliability seems to be somewhat larger for telephone in comparison to web data collection (Kogovšek 2006). Studies of the quality of measurement of network compositions (i.e., percentages of family, and friends in a network, Kogovšek et al. 2013) point to better measurement by face-to-face than by telephone.

In a study of reliability of measurement of whole networks across response scale formats (Ferligoj and Hlebec 1999), differences in validity were relatively small. Five-point response scales (both with labels for all values and only for extreme values) were found to be the most reliable, followed by drawing a line, the binary scale being the least reliable. The order of presentation of the scales to respondents was also registered – higher reliability for the scale used at later measurements – the so-called learning effect (see also Kogovšek 2006).

Finally, reliability of ethnic-religious and political composition and density of friendship and co-member networks are relatively high but lower for gender composition (Marsden 1993). Reliability of the core network is higher than that of the distant network (e.g., Morgan et al. 1997).

Historical Background

Data quality of "traditional" survey data (such as attitudes and opinions) has been studied from the 1940s on (Cantril 1944). Research into the quality of network data has started in the 1970s with the series of the so-called BKS studies, named after the main authors of these studies, Bernard, Killworth, and Sailer (Killworth and Bernard 1976, 1979/1980; Bernard and Killworth 1977; Bernard et al. 1979/1980, 1982, 1985). The main finding of these studies was that people are generally very inaccurate in reporting on their past interactions with other people. Later studies (e.g., Freeman and Romney 1987; Freeman et al. 1987) confirmed this finding but also showed that, on the other hand, people remember long-term or typical patterns of interaction with other people rather well. From the 1990s on, there has been an increasing and general awareness of the importance of social network data quality, and a large number of different studies with different methodologies have been conducted.

Approaches to Data Quality Evaluation

Test-Retest

The test-retest method involves administering the same questionnaire to the same respondents after some time. This elapsed time should be short enough to prevent true changes in the networks and long enough to prevent respondents from remembering their previous answers. If the questionnaire is very long (about half an hour), it is even feasible to ask the questions twice at the beginning and the end of the questionnaire. Otherwise, subjects have to be reinterviewed, which may be done earlier or later depending on the temporal stability of the networks. Reliability, or in other words stability, of any tie or network characteristic is estimated as the Pearson correlation coefficient between both repeated measurements.

In a study of test-retest reliability related to mood, Hlebec and Ferligoj (2001) showed that the change in respondents' mood was negatively correlated with stability: the larger the variability of mood, the larger the variability of (in)stability of measurement. Good mood at the time of measurement decreased the probability of unstable measurement.

Regarding stability of measurement of measures of centrality and prominence, Zemljič and Hlebec (2005) showed that:

- Global centrality and prominence measures are more prone to measurement errors than local measures.
- In-measures are more stable than out-measures.
- Stability of measurement is higher with the combination of line production and five-point response scales in the questionnaires and lower with the combination of line drawing and binary scales.

Use of External Criteria

Accuracy is best appraised when there is an external gold standard of the network or tie characteristic we intend to measure (criterion-related validity). Otherwise, validity is best appraised when there are external variables which are conceptually related to these network or tie characteristics (construct validity).

Gold standards should be strongly and positively correlated to questionnaire measurements. For instance, in a study of networks of companies, there may be accounting records of which companies buy from each. A valid measure of centrality of the company network computed from the questionnaire has to be strongly and positively correlated to the same measure as computed from the accountancy data.

As regards the correlations with external variables, they should be at the least of the expected sign. For instance, the reported frequency of contact with another network member for discussing personal matters should be positively correlated to a rating of how close the respondent feels to the other member, or extroversion of actors should be positively related to the proportion of friends in their personal networks (e.g., Kogovšek et al. 2013).

Split-Ballot Experiments

Split-ballot experiments proceed by randomly splitting the sample into two or more subsamples, each of which is measured with a different method. Most commonly, subsamples receive different versions of the questions or different methods of data collection. Typically, the distributions of the network characteristics of interest are compared across subsamples, in order to find differences in mean values or in frequency distributions across methods. While these experiments provide evidence on the existence of differences in the results across methods, they usually cannot lead to concluding which method is best.

An example application is the study of the quality of alternative question formats to measure the composition of ego-centered personal support networks (Kogovšek and Hlebec 2008). A very simple way to evaluate composition in a social network is the so-called role relation approach: to ask an ordinary survey question where response categories are types of components (e.g., partner, parents, children, friends). A more demanding method is the name generator approach.

The respondent first lists the names of all alters with whom he or she has some sort of relationship and then is asked additional questions to classify components. The study concludes that close family (partner and parents) appears as a bigger component with the role relation method and distant family, friends, neighbors, and coworkers with the name generator method.

Qualitative Techniques

The preliminary versions of questionnaires are often reviewed by experts. The best results are obtained when both experts in questionnaire design and experts in the substantive topic participate. Problems which can be easily uncovered by experts include complex syntax, vague meanings, excessively technical language, faulty list of answer categories, low correspondence between the questions and the concept which is to be measured, or difficulty for respondents to know the answer.

A small sample of respondents may be subject to a cognitive interview, in order to appraise the difficulties in understanding the question and providing the answer. For instance, participants may be required to think aloud when answering, to restate the questions in their own words, or to provide their own definitions to key terms in the questions.

A small sample of interviewers and respondents may be subject to a field pretest, conducted in circumstances as similar as possible as the large-scale survey. Interviewers may be later debriefed for problems encountered, or the full interviewer-respondent interaction may be audio recorded with permission. An analysis of responses may also reveal failures to understand or follow the questionnaire instructions, questions with very low variances, questions with very large missing data rates, and so on.

An example of such a study is that conducted by Nadoh et al. (2004) who used the so-called think aloud interview (a type of a cognitive interview) for evaluating measurement with the Kahn and Antonucci graphical method of measuring a social network. The focus of the study was how different respondents define the criterion of closeness for placing their network members into the three concentric circles, the closest persons being named in the innermost circle and the more distant persons in the outer circles. Respondents revealed using three criteria:

- Type of tie (e.g., “I thought about my family, my best friend. . . ”)
- Quality of tie (e.g., “I cannot imagine my life without them”)
- Formal properties (e.g., “these are the people with whom I have regular contacts”)

Key Applications

The study of questionnaire data quality has two major applications.

The first is to assist the choice between alternative question formulations, response scales, data collection modes, question orderings, and other design variables, in order to maximize reliability, validity, and accuracy of the network data.

The second application is to gather useful information to correct the bias in statistical results arising from measurement error. The easiest case is the correction of correlation attenuation. An estimate of the true correlation between two variables can be obtained by dividing the observed Pearson correlation with the geometric mean of the two estimated reliabilities.

Future Directions

Well into the 2000s, more complex studies on reliability and validity of network data are being done with meta-analytical procedures or with advanced statistical measurement models (e.g., confirmatory factor analysis models) on both whole and ego-centered networks (e.g., Kogovšek and Ferligoj 2005; Coromina and Coenders 2006). Measurement models can also be extended to structural equation models, which estimate data quality and correct attenuation bias in one single step (Kogovšek et al. 2013).

Measurement models are also being tailorized to the social network case. Multilevel confirmatory factor analysis models accommodate the study of quality of individual ties and network averages simultaneously (Coromina and Coenders 2006). Extensions to measurement models are also being developed for the analysis of quality of network compositions (Coenders et al. 2011). Data collection by web questionnaires is also receiving increasing attention (Vehovar et al. 2008).

It is still difficult to provide an unanimous answer to the question how to measure social networks and their characteristics. Extant studies on quality of social network data are very different by their research designs, types of networks studied (ego-centered, whole), definitions of data quality indicators, and types of samples used. Ultimately, the decisions with regard to measuring networks with the highest possible quality should be made in the context of an actual research design (type of network, study content, planned data analysis methods, etc.) and practical constraints (time, financial).

Cross-References

- ► Network Data Collected via the Web
- ► Questionnaires for Measuring Social Network Contacts

References

Bernard HR, Killworth PD (1977) Informant accuracy in social network data II. Hum Commun Res 4:3–18

Bernard HR, Killworth PD, Sailer L (1979/1980) Informant accuracy in social network data IV: a comparison of clique-level structure in behavioral and cognitive network data. Soc Netw 2:191–218

Bernard HR, Killworth PD, Sailer L (1982) Informant accuracy in social network data V: an experimental attempt to predict actual communication from recall data. Soc Sci Res 11:30–66

Bernard HR, Killworth PD, Kronenfeld D, Sailer L (1985) On the validity of retrospective data. Annu Rev Anthropol 13:495–517. Palo Alto, Stanford University Press

Bondonio D (1998) Predictors of accuracy in perceiving informal social networks. Soc Netw 20:301–330

Cantril H (1944) Gauging public opinion. Princeton University Press, Princeton

Casciaro T (1998) Seeing things clearly: social structure, personality and accuracy in social network perception. Soc Netw 20:331–351

Coenders G, Hlebec V, Kogovšek T (2011) Measurement quality in indicators of compositions: a compositional multitrait-multimethod approach. Surv Res Methods 5:63–74

Coromina L, Coenders G (2006) Reliability and validity of egocentered network data collected via web: a meta-analysis of multilevel multitrait multimethod studies. Soc Netw 28:209–231

Ferligoj A, Hlebec V (1999) Evaluation of social network measurement instruments. Soc Netw 21:111–130

Freeman LC, Romney AK (1987) Words, deeds and social structure: a preliminary study of the reliability of informants. Hum Organ 46:330–334

Freeman LC, Romney AK, Freeman SC (1987) Cognitive structure and informant accuracy. Am Anthropol 89:311–325

Hlebec V, Ferligoj A (2001) Respondent mood and the instability of survey network measurements. Soc Netw 23:125–140

Killworth PD, Bernard HR (1976) Informant accuracy in social network data. Hum Organ 3:269–286

Killworth PD, Bernard HR (1979/1980) Informant accuracy in social network data III: a comparison of triadic structure in behavioral and cognitive data. Soc Netw 2:10–46

Kogovšek T (2006) Reliability and validity of measuring social support networks by the web and telephone. Metodološki Zvezki 3:239–252

Kogovšek T, Ferligoj A (2005) Effects on reliability and validity of egocentered network measurements. Soc Netw 27:205–229

Kogovšek T, Hlebec V (2008) Measuring ego-centered social networks: do cheaper methods with low respondent burden provide good estimates of network composition? Metodološki Zvezki 5:127–143

Kogovšek T, Ferligoj A, Saris WE, Coenders G (2002) Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. Soc Netw 24:1–20

Kogovšek T, Coenders G, Hlebec V (2013) Predictors and outcomes of social network compositions: a compositional structural equation modeling approach. Soc Netw. 35:1–10 doi:10.1016/j.socnet.2012.10.002, http://dx.doi.org/10.1016/j.socnet.2012.10.002

Krackhardt D (1987) Cognitive social structures. Soc Netw 9:109–134

Lozar Manfreda K, Koren G, Vehovar V, Hlebec V (2004a) Collecting data on ego-centered social networks on the web: methodological issues. In: SUNBELT XXIV international social network conference, Portorož, 12–16 May 2004, p 162

Lozar Manfreda K, Vehovar V, Hlebec V (2004b) Collecting ego-centered network data via the web. Metodološki zvezki 1:295–321

Marsden PV (1993) The reliability of network density and composition measures. Soc Netw 15:399–421

Morgan DL, Neal MB, Carder P (1997) The stability of core and peripheral networks over time. Soc Netw 19:9–25

Nadoh J, Podreberšek P, Hlebec V (2004) Cognitive evaluation of the hierarchical approach for measuring ego-centered social networks. Metodološki Zvezki 1: 379–393
Pattison P (1994) Social cognition in context: some applications of social network analysis. In: Wasserman S, Galaskiewicz J (eds) Advances in social network analysis: research in the social and behavioral sciences. Sage, Thousand Oaks, pp 79–109
Vaux A (1988) Social support, theory, research, and intervention. Praeger, New York
Vehovar V, Lozar Manfreda K, Koren G, Hlebec V (2008) Measuring ego-centered social networks on the web: questionnaire design issues. Soc Netw 30:213–222
Zemljič B, Hlebec V (2005) Reliability of measures of centrality and prominence. Soc Netw 27:73–88

Recommended Reading

Carmines EG, Zeller RA (1979) Reliability and validity assessment. Sage, Beverly Hills
Groves RM, Fowler FJ Jr, Couper MP, Lepkowski JM, Singer E, Tourangeau R (2009) Survey methodology. Wiley, New York
Schuman H, Presser S (1996) Questions and answers in attitude surveys: experiments on question form, order and context. Sage, Thousand Oaks
Sudman S, Bradburn NM, Schwarz N (1996) Thinking about answers: the application of cognitive processes to survey methodology. Jossey-Bass, San Francisco

Quantitative Measurement

► Questionnaires for Measuring Social Network Contacts

Quantitative Political Science

► Legislative Prediction with Political and Social Network Analysis

Query Answering

► Query Answering in the Semantic Social Web: An Argumentation-Based Approach

Query Answering in the Semantic Social Web: An Argumentation-Based Approach

Maria Vanina Martinez[1] and
Sebastian Gottifredi[2]
[1]Department of Computer Science, University of Oxford, Oxford, UK
[2]Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina

Synonyms

Common sense reasoning; Decision making; Interaction network; Knowledge discovery; Query answering; Semantic Web search

Glossary

Web The World Wide Web
Semantic Web A collaborative movement that promotes common data formats on the World Wide Web and encourages the inclusion of semantic content in Web pages
Social Network Services (SNSs) Online services or sites that facilitate the building of social networks and relations among people who share interests, activities, backgrounds, or real-life connections
Social Web A set of social networks and relations that link people through the Web
Semantic Search Web search methods that seek to improve search accuracy by understanding the user's intentions and the contextual meaning of terms as they appear in the data space

Definition

The amount of information on the Web and the number of its human users have been growing exponentially in recent years. For many people, the Web has started to play a fundamental role as a means of providing and searching for information and services. The next revolution in Web search as one of the key technologies of the Web has just started with the incorporation of ideas from the Semantic Web, aiming at transforming current Web search into some form of semantic search and query answering on the Web, by adding meaning to Web contents and queries in the form of an underlying ontology. On the other hand, the Web has fostered the proliferation of Web-based communities and online social networks. Social networking services (SNSs) allow Web users to create and maintain an online network of close friends or business associates for social and professional reasons. The *Social Web* includes services such as Flickr, Facebook, Twitter, and last.fm, among many other social and business-oriented sites. This web of social sites, containing huge amounts of data and collective knowledge, demands the need for information integration, dissemination, reuse, searchability, and other more complex tasks such as querying and recommendation. The various and rich Semantic Web methodologies and tools seem to be the ideal platform to represent people and the objects that link them together in a social network. Furthermore, the use of standard representation formats facilitates the communication and linkage of data across different sites.

The application of Semantic Web methodologies to the Social Web is leading to a *Social Semantic Web*, a network of interlinked and semantically rich knowledge. The focus of research in the Social Semantic Web currently aims to the interlinking of documents, data, and even applications created by the end users themselves as the result of various social interactions, modeled and represented using machine-readable formats. In light of this vision, we propose a framework for query answering on the Social Semantic Web. Similar in spirit to the recent initiative from Google, called The knowledge graph (2013), we propose to leverage the complex knowledge base that emerges from the combination of social sites, networks, and applications to allow a user to query and extract general information from it. In a sense, we envision this framework to be used as a combination of semantic search facilitator and a recommender system, exploiting the rich and vast knowledge base that we, as users of the Web, feed with constant information about ourselves and the world that surrounds us.

Introduction

The Web has become a place where people create and maintain an online network of friends or business associates for personal and professional reasons through several social sites such as social networking services (SNSs). As we mentioned before, the combination of advancement in Semantic Web technologies and the proliferation of SNSs is creating the so-called Social Semantic Web, the interconnection and interoperation of social sites within the Web. Such integration allows for the creation of a richer model of social networks, one that considers, apart from the structure of the network, the shared objects which bring people together, as well as the semantics that can be used to represent these interlinked people and objects. In general, Web users participate in multiple disconnected or loosely connected and heterogeneous social sites. Just to give a simple example, it is very common for a Web user to have an account on Facebook, one on Twitter, use Flickr or similar sites to post photos, and also use a social bookmarking site such as Delicious. If we consider questions such as *What new book should I read now?*, *Which bar in this area could be a good place to go and relax for half an hour?*, *What present could I buy for my friend?*, or *In what things is my friend interested?*, the conglomeration of social networks to which users adhere contains rich and (potentially) useful information that can help in answering this type of queries. By means of their connections and behavior within the Web, people implicitly, and sometimes explicitly, define their profile and declare their tastes, preferences, opinions, etc.

In this work we propose a framework for query answering in the Semantic Social Web that connects the various SNSs and use this as a complex knowledge base centered on a user that can query and extract relevant information. We formalize our framework on the premise that it is possible to use the research developed both in Semantic

Web and Social Network analysis, as well as that from ontology languages, natural language processing, information retrieval, filtering systems, opinion analysis, etc. The combination of these foundational methods and tools allows us to put together a comprehensive knowledge base and to perform complex reasoning over it.

A very important issue that is bound to arise as the result of the integration of the contents of such social networks is that of conflicting information. Though the adoption of a standardization under a common semantics potentially reduces the amount of contradictory information that comes from misrepresentation of concepts and relationships, it is almost inevitable, and sometimes expected or even desired, to encounter conflicting information in a complex and useful knowledge base. Each social network models a different aspect of a user, and though characteristics and relations might overlap among networks, it is reasonable to assume that each model is incomplete and potentially inconsistent with respect to the rest. How, then, do we make sense of this highly dynamic and (potentially) contradictory information? How can we justify the answers to a query? Argumentation represents a sophisticated mechanism for the formalization of commonsense reasoning, which has shown its importance in different areas of Artificial Intelligence (AI) such as multi-agent systems, decision support systems, and legal systems among others (see Rahwan and Simari 2009). Intuitively, an argument is a coherent set of information that supports a claim. The acceptance of this claim will depend on a dialectical analysis (formalized through a proof procedure) of the arguments in favor of and against the claim (Rahwan and Simari 2009). Argumentation provides a natural way of reasoning with conflicting information, which retains much of the process that a human being would apply in such situations. Furthermore, though the reasoning process is automated, it does not have to be completely transparent to the user. Argumentation frameworks have the added value of providing an explanation for the results of the reasoning process, which allows a user to analyze and understand where the answers come from.

Background and Related Areas

The Web has dramatically changed since the beginning of the Semantic Web effort, which aims to add well-defined meaning to objects and better enables computers and people to work in cooperation. In the last decade, many important advances have been made towards the definition of the foundational standards supporting data interchange and interoperation that allow today the creation of metadata and associated vocabularies to semantically enrich Web sites and Web-related applications. The use of such technologies allows to build complex knowledge bases that are interconnected through a common semantics. A challenge that arises from this is the problem of reasoning over that data, e.g., how to combine knowledge that is distributed over many Web pages to answer complex queries or perform recommendations to Web users. The area of *ontological query answering* has produced many interesting and promising results in this direction, both for description logics (Zhou et al. 2012) and rule-based ontological knowledge bases (Fazzinga et al. 2010; Mugnier 2011).

Extensive work has also been carried out in the area of *semantic recommender systems*. Traditionally, recommender systems, or information filtering systems, have been developed to apply information filtering techniques that allow the user to access specific information that he needs. Most of these systems assist the user in information retrieval tasks (Adomavicius and Tuzhilin 2005; Peis et al. 2008) or are used to predict the user's valuation on not yet evaluated items. In general, they are domain specific: news, emails, scientific information services, music, movies, e-commerce, etc. Different filtering methods have been developed, and they can be classified, for instance, by the type of information used to perform the filtering. *Content-based filtering* systems generate recommendations using the user's preferences. On the other hand, *social filtering* systems use the information associated with users with *similar* characteristics. Most recently, semantic approaches to recommendations have been developed. These systems are mainly designed as a knowledge base normally defined through an ontology or conceptual map, and that use technologies from the Semantic Web to create the recommendations. Most of these systems are based on users' profiles containing long-term information, needs, preferences, interests, etc. Recommendations are created based on the semantic distance or similarities among users. How these distances are computed depends heavily on the application and the particular approach. In general, a combination of content-based filtering and social filtering is performed, enhanced by the knowledge contained within the underlying ontology.

Nowadays, the argumentation community is starting to develop approaches that handle some of the problems identified in the Social Web. These research lines are mainly focused on providing support through argumentation to the informal debates and dialogues occurring in this context (Toni 2012; Heras et al. 2010). For instance, in Toni (2012) the author outlines a methodology for analyzing the informal exchanges (e.g., debates or discussions) in the online social platforms in terms of computational argumentation. This approach identifies *comments*, *opinions*, and *links* in the informal debate and translates this into *abstract argumentation frameworks*. Using these frameworks, a standard argumentative analysis can be carried out to provide a formal assessment of the dialectical validity of the positions being debated.

Finally, the *Social Semantic Web* was presented in Breslin et al. (2009) as the future of the Web as a "network of interlinked and semantically-rich content and knowledge." The authors describe some popular social media and social networking applications and show an approach for interconnecting the social Web sites with semantic technologies and for enhancing semantic applications with content created from those communities. There are already many efforts in the Semantic Web to augment the structure, the usage, and the combination of social networking and social Web sites. For instance, the Friend-of-a-Friend (FOAF) project (2013) allows to describe people and relationships, Nepomuk social semantic desktop (2008) is a framework for extending the desktop to a collaborative environment for information management and sharing, and the Semantically-Interlinked Online Communities (SIOC) initiative (Breslin et al. 2005) represents online discussions; some social networking services, such as FriendFeed (2013), are also starting to provide query interfaces to their data and allow interoperation between other social sites such as Facebook, Flickr, Twitter, etc. The Social Semantic Web is a very ambitious project that is recently starting to see results. Nevertheless, in this work we frame a proposal to advance the state of the art one step further by addressing the problem of obtaining semantically rich answers from the complex knowledge base that Web represents.

Proposed Solution and Methodology

We start by defining a model of the Social Web that is centered on a user. A social network represents an abstraction of social structures that link individuals, objects, and organizations. We can think of a social network as a dynamic graph with the following characteristics:

- *Nodes* represent objects in the network, which represent either individuals, objects or products, locations, or organizations. Each node is a typed complex structure that contains a profile of the entity that it represents.
- Nodes are connected through *ties* or *links*. The format of these links is rich, and they can also be typed since they represent the nature of connections between nodes, which can vary depending on the purpose of the network.

The Social Semantic Web is then the conglomeration of the social networks to which a user belongs; projects such as FOAF and ontologies such as SIOC (Breslin et al. 2005) provide adequate tools to model these entities and semantically interlink them. Instead of considering all social networks to which a user belongs as a unique interlinked *machine-understandable graph layer* (with nodes as users or related data, and arcs as relationships) over the existing Web of documents and hyperlinks (Breslin et al. 2009), we assume that we can also identify each network as an individual *information module*.

An information module $\mathcal{M}$ is a graph (V, E) where each node $v \in V$ is composed of an identifier *id* that is unique in the graph, and there is a type t identifying if the entity represented by the node is either an individual, an object, a location, or an organization, a profile $P(v)$, and a log $L(v)$. Profile $P(v)$ contains demographic information about the entity represented by the node, explicit preferences and tastes, interests, and goals; log $L(v)$ contains a record of activities that describe the entity's behavior within the network. Formally, a profile is a set of pairs $\langle propertyID, value \rangle$ where *propertyID* represents a property of the object, and *value* represents the description of that property for that particular node. The log stores activities in the form of a time stamp and a description of the activity. Note that the actual contents of the profiles and activity logs may respond to the underlying ontology used to represent the networks.

Each edge $e \in E$ has the form $e = (v_i, v_j, desc)$, where $v_i, v_j \in V$ and *desc* characterizes the type of link between the nodes; this characterization includes the directionality of the link, the meaning and purpose of the link, and how it relates to other types of links; for this we assume that the underlying ontology provides a way of characterizing complex ties among objects.

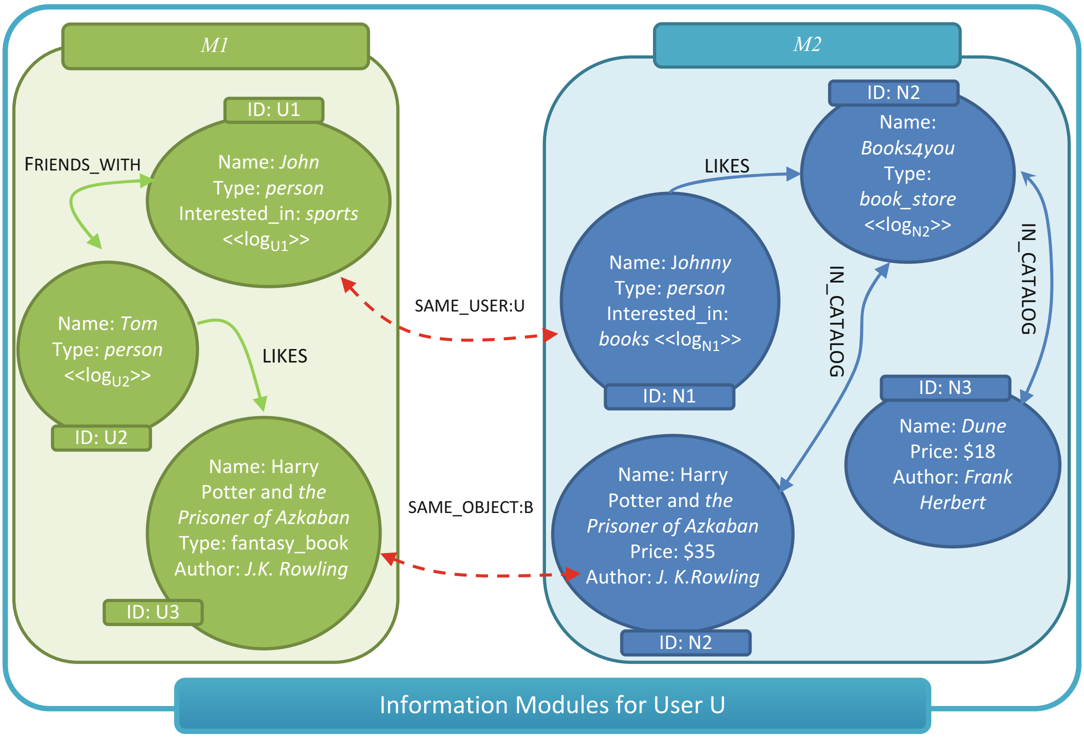
The framework we propose is centered on a user, represented as a set of nodes within the different networks, and reasoning is performed from the point of view of this user. Note that we assume that the underlying ontology used to represent the social networks provides a way to identify that several nodes in different (or the same) networks represent the same entity. We identify then the *Personal Semantic Social Web* for a given user. A user is a set of nodes from different networks. The user's personal semantic social Web is then a structure containing (1) a set of information modules $\mathcal{M}_1, \ldots, \mathcal{M}_n$, representing the social networks to which the user belongs; (2) the set of nodes corresponding to the user from each information module, $V_U = \{v_1, \ldots, v_k\}$ with $v_i \in \bigcup_{M_i = \langle V, G \rangle} V$; (3) the user's global profile $\mathcal{P}_U = \bigcup_{v \in V_U} P(v)$ that combines the user's different profiles; and (4) the user's global activity log $\mathcal{L} = \bigcup_{v \in V_U} L(v)$.

Example 1 Figure 1 shows the Personal Semantic Social Web for user U. The user is identified by node U_1 in module $\mathcal{M}_1$ and node N_1 in module $\mathcal{M}_2$. The global profile of user U is $\mathcal{P}(U) = \{\langle type, person \rangle, \langle name, John \rangle, \langle name, Johnny \rangle, \langle interested\_in, books \rangle, \langle interested\_in, sports \rangle\}$; the global activity log of user U is $\mathcal{L} = log_{U_1} \cup log_{N_1}$.

Additional semantic information in the modules shows that user U *has a friend* in $\mathcal{M}_1$ (node U_2) and that he *likes* node N_2 in module $\mathcal{M}_2$. Node N_2 represents a book store and $\mathcal{M}_2$ also contains the books that belong to the book store's catalog.

There has been a tremendous amount of work on the analysis of social networks. These works provide different models to qualify and quantify relationships within the networks, to measure particular properties or phenomena such as centrality/influence of nodes (Freeman 1978; Borgatti and Everett 2006; Freeman 1977) and diffusion or propagation of information (Bröcheler et al. 2010; Cha et al. 2009; Jackson and Yariv 2005; Kang et al. 2012). On the other hand, query answering on social networks can often be described as a subgraph matching problem, i.e., the query can be represented as a subgraph, with partially instantiated properties, and, in order to answer such query, a process that involves matching the query subgraph to a subgraph in the network is required. This issue has been studied extensively in both the Semantic Web community (Harth et al. 2007) and the social network communities (Bröcheler et al. 2009), with very promising results regarding efficiency and quality of answers (Ronen and Shmueli 2009; Bröcheler et al. 2009, 2011).

The queries that we aim to answer within the Personal Semantic Social Web of a user involve further use and understanding of the knowledge that arises from the collection of the different social networks. The combination of the work in social network analysis and query answering



Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 1 Personal Semantic Social Web for user U. Information modules M_1 and M_2 represent two different social networks to which the user belongs; user U is identified by node U_1 in module M_1 and by node N_1 in module M_2

in social networks within our framework allows to envision the possibility of answering complex queries that require a combination of node matching and the analysis of the networks' structure.

With the objective of maintaining this work as general as possible, we do not adopt any particular query language; instead, we describe the general form of queries and answers showing the information they should convey.

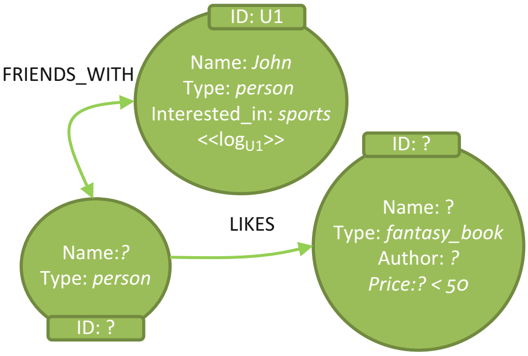
Definition 1 Let U be a user; a query from U specifies (1) a partial profile $\mathcal{P}_Q$, i.e., a set of pairs $\langle propertyID, value\rangle$ where *value* could be left unspecified, and (2) a list of constraints $\mathcal{C}_Q$ that an answer must satisfy. We use the notation $Q = \langle \mathcal{P}_Q, \mathcal{C}_Q \rangle$.

Example 2 Consider user U from Example 1; suppose he wants to obtain from his Personal Semantic Social Web the following information: *the recommendation of fantasy books that his friends like and cost less than $50*. Independently of the query language adopted, a query Q is created such that $\mathcal{P}_Q = \{\langle type, book\rangle, \langle genre, fantasy\rangle\}$. The set of constraints imposed by the query is $\mathcal{C}_Q = \{c_1 : price \leq 50,\ c_2 : \text{“}my\ friends\ like\text{”}\}$.

Definition 2 Let $Q = \langle \mathcal{P}_Q, \mathcal{C}_Q \rangle$ be a query issued by user U. An *answer* to Q is a node n in V_i for some $(V_i, E_i) = \mathcal{M}_i$ where M_i is an information module for user U, such that $\mathcal{P}(n)$ matches $\mathcal{P}_Q$ and n satisfies the constraints imposed by Q.

Example 3 Looking at Fig. 1, we can see that an answer to query Q from Example 2 is a node that represents the book *Harry Potter and the Prisoner of Azkaban* by *J. K. Rowling*; since its genre is fantasy, the price is between the range specified by c_1, and a user's friend, namely *Tom*, likes it.

In this work we do not focus on the details of how to obtain the answers to a query; instead,



Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 2 Query Q from Example 2 represented as a graph

we assume that this process is carried out by an adequate method chosen in accordance with the underlying representation of the networks. For instance, query Q from Example 2 could be represented as the graph in Fig. 2. The process of finding the answers to Q then involves matching the subgraph to subgraphs in the information modules. Works such as Ronen and Shmueli (2009) and Bröcheler et al. (2009) provide efficient mechanisms for specific representation languages (Datalog and RDF, respectively).

Argumentation-Based Query Answering in Social Networks

In the previous section we defined an answer to a query as any node such that its profile matches the query's partial profile and it satisfies the constraints imposed by the query. However, not all answers are equally useful. The recommendation aspect of the query implicitly suggests that the user is looking for books that *he might want to read*; then, other information in the different modules that imply that *Tom* is not a *book-compatible friend* for user U should be considered. Another implicit requirement in the query is that even though the user only asked for books that his *friends* like, he wants only his *most relevant* friends to be taken into account. In this work we pursue the identification of answers that are *guaranteed* to be relevant for the user. In order to do this, it is necessary to consider all evidence for and against each particular answer; we thus propose an argumentation-based approach to query answering.

Argumentation is a form of reasoning where a claim is accepted or rejected according to the analysis of the arguments for and against it. The way in which arguments and justifications for a claim are analyzed provides an automatic reasoning mechanism for knowledge bases with contradictory, incomplete, dynamic, and uncertain information (Rahwan and Simari 2009). In the last decades argumentation has evolved as an attractive paradigm for conceptualizing commonsense reasoning, and its study within the field of Artificial Intelligence has been growing steadily (Bench-Capon and Dunne 2007). Several approaches were proposed to model argumentation on an abstract basis (Dung 1995), using classical logics (Besnard and Hunter 2001) or using logic programming (García and Simari 2004; Prakken 2010).

The process of query answering then relies on an argumentative analysis of the information contained in the user's information modules. Informally, first potential answers are computed, i.e., the nodes that match and satisfy the query, and each potential answer is associated with an argument that supports its relevance. An exhaustive analysis of conflicting information is performed for each such argument, resulting in the decision of whether or not a potential answer is returned to the user as a guaranteed answer; this happens if and only if the evidence in favor of its relevance outweighs the evidence against it. In the following, we first introduce the argumentation framework that we adopt in this work, and we then go on to show the specifics of how the different elements for performing query answering in this setting are instantiated in such a framework and how guaranteed answers are obtained.

Typed Argumentation Basics

In this work we base our research on abstract argumentation (Dung 1995). That is, here, arguments are abstract entities that will be denoted using uppercase letters, and no reference to the underlying logic is needed since we are abstracting its structure. We abstract away how the arguments are built as well as how the relations among them (conflicts, preferences, and defeats) are obtained. Dung's work on abstract argumentation has been seminal for the study of argumentation and has played a major role as a way of understanding argument-based inference. Moreover, this kind of frameworks is of particular interest because they can be instantiated with concrete languages and mechanisms to construct arguments and determine defeats. In particular, we will use a variation of Dung's abstract argumentation framework (Dung 1995) that explicitly considers the notion of typed arguments. The idea of argument type involves the characterization of a group of arguments that have some common features and, therefore, these features distinguish them from other groups of arguments.

In the context of abstract argumentation, the concept of argument type was formalized in Multi-Typed Argumentation Frameworks (MTAF) (Gottifredi et al. 2011). Single-Typed Argumentation Frameworks (STAFs) can be used to represent individual argument types and to characterize special features (e.g., conflict and preferences) of the group of arguments they represent.

Definition 3 A Single-Typed Argumentation Framework (STAF) is a tuple $ST = (AR, \rightsquigarrow, \succeq)$, where AR is a finite set of arguments, $\rightsquigarrow$ is an attack relation between arguments of AR, and $\succeq$ is an anti-reflexive preference relation over the arguments of AR.

The functions $\mathsf{Args}(ST)$, $\mathsf{Atts}(ST)$, *and* $\mathsf{Prefs}(ST)$ will respectively return the arguments, the attacks, and the preferences of ST.

Consider a set of arguments $\{\mathsf{A}_1 \dots \mathsf{A}_n\}$ such that every argument in this set shares certain features and therefore are considered as arguments of type T_A. Let us also consider that there is another set of arguments $S_B = \{\mathsf{B}_1 \dots \mathsf{B}_k\}$ of type T_B and, in the modeled domain, arguments of T_A are always preferable to arguments of type T_B. Both S_A and S_B can be modeled by two different STAFs. However, suppose there is also a specialization of T_B (called T_C), that is, a subset $S_C \subset S_B$ that inherits some properties of T_B, but all arguments of type T_C are preferable to arguments of type T_A. A more complex structure is needed to represent such scenarios. We then introduce MTAFs, which contain several STAFs (representing every type in the system), and allow to relate these individual types through inheritance, preference, or conflicts.

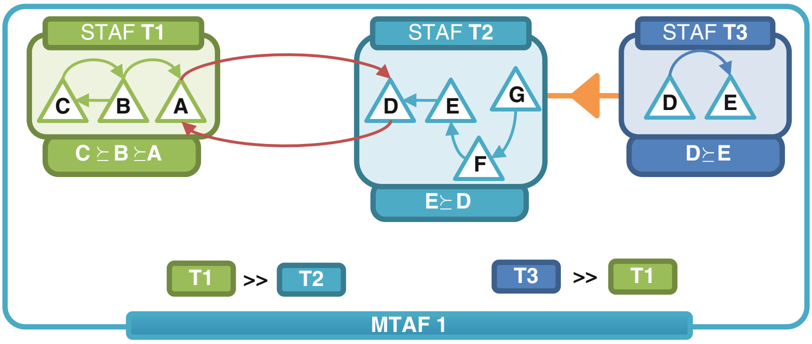
Definition 4 A Multi-Typed Argumentation Framework (MTAF) is a tuple $MT = (S_T, -\!\!\triangleright\!\!-, \mathsf{TAtt}, >>)$, where:

- S_T is a set of STAFs.
- $-\!\!\triangleright\!\!-$ is the inheritance relation, a reflexive and transitive relation over S_T such that if $(T_i, T_j) \in -\!\!\triangleright\!\!-$ then it holds that $\mathsf{Args}(T_i) \subseteq \mathsf{Args}(T_j)$.
- TAtt is the type attack function defined over $S_T \times S_T$ that returns the attacks among the arguments of different types.
- $>>$ is the type preference relation, which is an anti-reflexive relation over S_T.

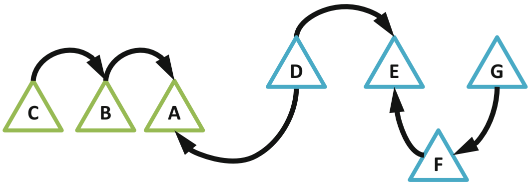
The function $\mathsf{MArgs}(MT)$ will return the arguments of all the STAFs in MT, that is, $\mathsf{MArgs}(MT) = \bigcup \mathsf{Args}(T_i), T_i \in S_T$.

Example 4 Consider MTAF 1 depicted in Fig. 3. There it can be seen that this MTAF has three individual types characterized by the STAFs $T1$, $T2$, and $T3$. Each STAF depicts its arguments as triangles, attacks among them with bold arrows, and preferences among them with the $\succeq$ symbol. For instance, $T1$ contains the arguments A, B, and C; it denotes that C and B attack each other and that B attacks A, and it establishes that C is preferred to B and that B is preferred to A. Regarding the type relations, it can be seen that $T3$ inherits from $T2$, there is a mutual attack between the arguments A of $T1$ and the argument D of $T2$ (and $T3$), the type $T1$ is preferred to the type $T2$, and the type $T3$ is preferred to the type $T1$.

In argumentation, the challenge consists in finding out, all things considered, which arguments prevail, i.e., those arguments that are accepted. For this it is necessary to analyze the defeat relation among the framework's arguments. When an argument A defeats an argument B, it means that A and B are in conflict (i.e., A attacks B) and A is as preferred as B (Amgoud and Cayrol 2002). That is, a defeat



Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 3 An example of an MTAF



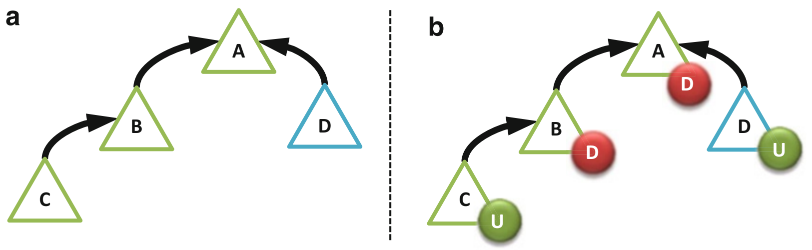
Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 4 Global defeats

can be seen as an effective attack. In multi-typed argumentation, to determine the defeat relation, we have to consider conflicts and preferences among arguments of each type, conflicts and preferences among types, and the inheritance relation among types. As shown in Gottifredi et al. (2011), all these relations are considered in the construction of a *global defeat relation* among all the arguments of the MTAF. The intuition behind this global defeat relation resides in the notion of typespecialization; there are relations (preferences and conflicts) for more specialized types that may override relations for less specialized types. For more details on how global defeats are obtained, refer to Gottifredi et al. (2011). Next, in Fig. 4, we show the global defeat relation for the arguments of the MTAF of Example 4.

Using the defeat relation it is possible to determine whether or not an argument is accepted. To this end, the notion of *argumentation semantics* has been extensively studied (Baroni and Giacomin 2007). In this work, the acceptance of an argument will depend on a dialectical analysis, which provides a useful way of characterizing proof procedures for skeptical semantics in argumentation (Prakken 2005). These dialectical proof procedures have been recognized in the literature for their usefulness in query-answering scenarios (García and Simari 2004).

The dialectical analysis models defeasible reasoning as a dispute between two parties (Proponent and Opponent) who exchange arguments and counterarguments. An Argument A is provably justified if every *defeater* presented by the Opponent can be ultimately defeated by the Proponent. Each of these dialogues is usually called an *argumentation line*. In our approach, each line is a nonempty sequence λ of arguments from an MTAF, where each argument in λ defeats its predecessor in the line. An argumentation line should be noncircular (an argument should not occur twice in the same argumentation line) in order to avoid infinite lines, and it should also be exhaustive, i.e., no more arguments can be added to it.

Definition 5 Given an MTAF MT, where $\rightarrow$ is the global defeat relation among MArgs (MT) and $\mathsf{B}_1, \ldots, \mathsf{B}_n \in \mathsf{MArgs}(MT)$, an **argumentation line** λ in MT is a (nonempty) finite sequence of arguments $[\mathsf{B}_1, \ldots, \mathsf{B}_n]$ such that $\forall \mathsf{B}_i, \mathsf{B}_j$ with $i \neq j$ and $1 < i, j \leq n$, $\mathsf{B}_i \rightarrow \mathsf{B}_{i-1}$, $\mathsf{B}_i \neq \mathsf{B}_j$ and $\nexists \mathsf{C} \in \mathsf{MArgs}(MT)$ such that $\mathsf{C} \rightarrow \mathsf{B}_n$. The argumentation line λ is said to be rooted in B_1. The set of all argumentation lines in MT is noted as $MT_{\lambda s}$.



Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 5 (**a**) Dialectical tree for A. (**b**) Marked dialectical tree for A

The set of all possible argumentation lines leads to a tree structure called dialectical tree (García and Simari 2004).

Definition 6 Given an MTAF MT and an argument $\mathsf{A} \in \mathsf{MArgs}(MT)$, the **dialectical tree** $\mathbb{T}(\mathsf{A})$ rooted in A from MT is built from the set X of every argumentation line of MT rooted in A, and is such that an argument C in $\mathbb{T}(\mathsf{A})$ is:

- A node iff A appears in a line $\lambda \in X$
- A child of a node B in $\mathbb{T}(\mathsf{A})$ in λ iff $\lambda = [\ldots, \mathsf{B}, \mathsf{C}, \ldots]$, $\lambda \in X$
- A leaf of $\mathbb{T}(\mathsf{A})$ in λ iff C is a leaf in $\lambda \in X$

In a dialectical tree, every node can be marked as defeated (**D**) or undefeated (**U**); leaves are marked **U** and the inner nodes, including the root, are marked **D** when they have at least one child marked **U** or marked **U** when all their children are marked **D**. A marked dialectical tree represents a dialectical analysis considering every possible argument that can be used for and against the argument in the root of the tree. Therefore, if the root argument is marked undefeated, it means that this argument can be considered as accepted. Next, in Fig. 5a we depict the dialectical tree for the argument A of the MTAF of Example 4, and in Fig. 5b we show its marked dialectical tree.

Computing Arguments and Counterarguments for Query Answering

The first step towards computing *guaranteed answers* for a query is to compute the set of its answers, as was previously shown. Given a query Q issued by user U, we will call the set of answers to Q *potential* answers. We associate with each potential answer p a set of arguments $Args_p$. Each argument A_{i_p} in a $Args_p$ represents a support for p as an answer to Q. We need now to consider every other piece of information that might conflict with this support.

Example 5 Consider again the information modules for user U in our running example. As shown in Example 3, the node corresponding to the book *Harry Potter and the Prisoner of Azkaban* is a potential answer for Q from Example 2; we will call this potential answer p. We associate an argument A_p to p. This argument supports p as an answer for Q given that p satisfies $\mathcal{P}_Q$ and *Tom*, a friend of user U, likes that book. Note that a different argument could be associated with p if, for instance, another friend of user U, within information module M_1 or from another information module, also likes the same book.

We use the multi-type argumentation framework defined in the previous section. We associate a STAF ST_{M_i} with each information module M_i for a user U; therefore, all arguments based on knowledge from module M_i belong to ST_{M_i}. Preference relations among arguments within a STAF can be informed by explicit information stated by the user, by the application of different analysis techniques to the module, or by the combination of explicit and mined information. As an example, approaches such as sentiment or opinion analysis (Subrahmanian and Reforgiato Recupero 2008; Pak and Paroubek 2010) can provide confidence values on the strength of the arguments. For instance, suppose that an argument supporting the conclusion that a user likes Harry Potter's books is solely based on the fact he once wrote the comment "*I'm really enjoying the new Harry Potter book*," while another argument can be based on a series of detailed and eloquent comments about the literary quality of the Harry Potter series. It seems that the second argument should be preferable to the first one. Another example of tools that can help in defining preferences among arguments is statistical analyses, measuring frequency of certain actions within the networks such as *likes*, *tags*, *shares*, and *retweet*. Finally, properties of the nodes in the networks could also influence the preference among arguments they promote. For instance, arguments whose support comes from influential players in a network could be preferable to those promoted by others. There are many approaches in the literature that can be leveraged for this purpose, such as Borgatti and Everett (2006), Cha et al. (2009), Jackson and Yariv (2005), and Kang et al. (2012).

Definition 7 Given a STAF $ST_{M_i} = (AR, \rightsquigarrow, \succeq)$ and an information module M_i, we assume the existence of a function $\mathcal{F}$ that takes ST and M_i and returns $ST' = (AR', \rightsquigarrow', \succeq')$, a new STAF such that: (1) $\rightsquigarrow \subseteq \rightsquigarrow'$, (2) $\succeq \subseteq \succeq'$, and (3) $AR \subseteq AR'$ with $\mathsf{A} \in AR' \setminus AR$ iff there is $(\mathsf{A}, \mathsf{B}) \in \rightsquigarrow'$ where $\mathsf{B} \in AR$.

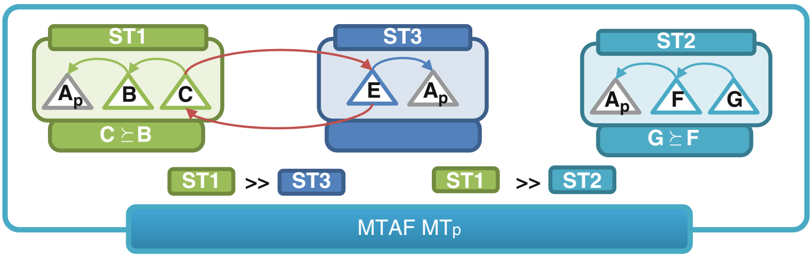
Function $\mathcal{F}$ *exstends* a STAF w.r.t. module M_i by both finding arguments based on the information contained in M_i that contradicts or is in conflict with arguments in the STAF and extending the preference relation. For instance, consider a potential answer p for a query about buying books. Suppose that p is a node representing a book from the Harry Potter series and that the corresponding argument in its support is A_p. Let $ST_{\mathcal{M}_i}$ be a STAF associated with module $\mathcal{M}_i$ that contains argument A_p. Suppose now that we can compile from $\mathcal{M}_i$ another piece of information c that yields evidence towards the fact that the user thinks that *Harry Potter is boring*. Clearly, in this scenario, the piece of information c is in conflict with p. Thus, considering this situation, the STAF $ST_{\mathcal{M}_i}$ should also contain an argument C supporting c and the attack $\mathsf{C} \rightsquigarrow \mathsf{A}_p$ denoting the conflict between c and p. This means that a STAF associated with module $\mathcal{M}_i$ contains the set of arguments supporting a potential answer and every other argument (from $\mathcal{M}_i$) that is in conflict with those arguments; we call these counterarguments. Furthermore, since we need to perform an exhaustive analysis of information in support and against the potential answer, the STAF also needs to consider arguments that are in conflict with the counterarguments, and those in conflict with the latter, and so on and so forth. We will see below how this is accomplished.

We can now define an MTAF based on which it is possible to determine whether or not a potential answer should be returned as a guaranteed answer. For every potential answer p to Q, we define an MTAF MT_p containing a STAF $ST_{\mathcal{M}_i}$ for each information module $\mathcal{M}_i$ for user U.

Given a potential answer p and corresponding arguments $Args_p$, the STAFs $ST_{\mathcal{M}_i} = (AR_i, \rightsquigarrow_i, \succeq_i)$ within the MTAF MT_p are obtained in the following way:

1. Initially, we have $ST^0_{\mathcal{M}_i} = (AR^0_i, \rightsquigarrow^0_i, \succeq^0_i)$ with $AR^0_i = Args_p$, $\rightsquigarrow^0_i = \emptyset$, and $\succeq^0_i = \emptyset$.
2. Then, $ST^j_{\mathcal{M}_i} = \mathcal{F}(ST^{j-1}_{\mathcal{M}_i}, M_i)$.
3. $ST_{\mathcal{M}_i} = ST^k_{\mathcal{M}_i}$ for some $k > 0$ such that $ST^k_{\mathcal{M}_i} = ST^{k-1}_{\mathcal{M}_i}$.

The preference relation among STAFs within MT_p can be defined also directly by the user or mined for instance from the analysis of his activity logs corresponding to each network. For instance, preferences can state that user U has more entry logs related to books in information module $\mathcal{M}_1$ than in module $\mathcal{M}_2$; therefore, for queries related to books, a STAF from $\mathcal{M}_1$ is preferred over one from $\mathcal{M}_2$. On the other hand, attacks among the arguments of the different STAFs within MT_p are represented by the TAtt function in MT_p. These attacks are determined by analyzing the information supported by arguments of the different STAFs in MT_p. Note that this is similar to the way in which the function $\mathcal{F}$ finds the arguments to build each STAF of MT_p. For instance, if in the STAF $ST_{\mathcal{M}_j}$ we have the argument D_s supporting a piece of information s that indicates that the user *hates magic and magic-related topics*, then function TAtt will establish that D_s attacks A_p.



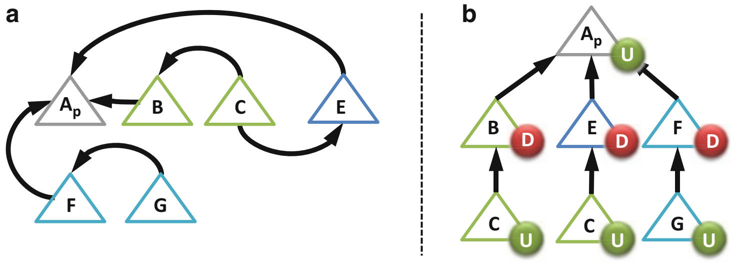
Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 6 MTAF MT_p for potential answer p to Q from Example 3

Example 6 In Fig. 6 we can see an MTAF for the potential answer p from Example 5. As it can be seen, there are three STAFs, $ST1$, $ST2$, and $ST3$, each of which corresponds to the information modules M_1, M_2, and M_3, respectively, from Fig. 1 (M_3 is not shown in Fig. 1 for simplicity). STAF $ST1$ contains three arguments: A_p supporting potential answer p, argument B attacking A_p, and argument C attacking B. The attack from B to A_p could state, for instance, that p is not a good answer because *Tom* and user U are not really friends, based on statistical information from module M_1 that shows evidence of the few and sparse interactions between the users in the network. On the other hand, the attack from C to B could state that *Tom*'s and the user's profiles are *very similar* and therefore it is highly probable that they like the same types of books, so p is actually a good candidate. In STAF $ST3$ there are two arguments: A_p and E. Argument E attacks A_p stating, for instance, that p is not a good answer since from module M_3 it is possible to collect information supporting the fact that user U does not like magic-related topics or products. Furthermore, argument E also attacks argument C from STAF $ST1$ (and vice versa). Finally, in STAF $ST2$, argument F attacks A_p; since M_2 contains information for a bookstore, this argument could attack A_p on the grounds that user U has not bought any of the Harry Potter books in the past, or any other book from the same author or editorial. In the same way, argument G could attack F since the user has bought plenty of fantasy books that are *similar* to the Harry Potter series. As stated before, preferences among arguments within a STAF or among STAFs can be extracted in many different ways. In this example, we have that argument C is preferred to B in $ST1$, that G is preferred to F, and that $ST1$ is preferred to $ST2$, and $ST1$ to $ST3$.

The following step is to analyze MTAF MT_p to determine the acceptability status of each A_{i_p}, that is the arguments supporting p in MT_p. For this, using MT_p, we will build a dialectical tree $\mathbb{T}(\mathsf{A}_{i_p})$ for each A_{i_p}. If at least one A_{i_p} is considered as accepted (A_{i_p} is marked $\mathbf{U}$ in $\mathbb{T}(\mathsf{A}_{i_p})$), then p will be guaranteed and returned as the answer of the query.

Example 7 In Fig. 7a we depict the global defeat relation that can be obtained from the MTAF MT_p from Example 6. In particular, note that argument C defeats argument E because, as mentioned in Example 6, C belongs to ST_1 that is preferred to ST_3, the STAF E belongs to. Using the defeat relation among these arguments, in Fig. 7b, we show the marked dialectical tree for the argument A_p. Observe that A_p is marked as $\mathbf{U}$, and therefore p will be guaranteed and returned as an answer for the query Q of Example 3.

One of the main advantages of argumentation-based frameworks over other types of decision-making approaches is the possibility of explaining to the user how, and why, conclusions are reached within the system. Dialectical trees $\mathbb{T}(\mathsf{A}_{i_p})$ can be used as explanations for why certain potential answers are returned to the user as guaranteed answers while others are rejected, showing clearly the roles played by the different pieces of (conflicting) information



Query Answering in the Semantic Social Web: An Argumentation-Based Approach, Fig. 7 (**a**) Defeat relation among the arguments of MT_p from Example 6. (**b**) The marked dialectical tree for the argument A_p of MT_p

during the query-answering process. This type of information can be further leveraged if the user has the possibility of providing feedback to the system; being aware of how the answers are obtained, the user could, for instance, adjust the preferences among information modules, or ask the system to personalize the usage of the methods and measures for analyzing the network structure, etc.

Key Applications

As presented here, this framework is envisioned to work as a query-answering/recommender system within the *Personal Semantic Social Web* of an individual, making the explicit and implicit information contained in his online communities available and producing semantically rich answers to complex queries, as well as high quality recommendations. A semantic Web search engine can be built within this framework to enrich and guide the search in the Web with the social environment of the user.

Another promising application, as proposed in Toni (2012), is to allow the user to issue a query asking for the status of discussions that are happening within his social networks. Argumentation-based frameworks are especially well suited for that type of dialogue-based reasoning. In addition, our framework would allow to analyze such discussions across networks.

Future Directions

In this work we provide a high-level description of an argumentation-based system for query answering in the *Personal Semantic Social Web* of a user. Future directions of this work include the instantiation of the framework using specific argumentative representation languages such as DeLP (García and Simari 2004) or ASPIC (Prakken 2010), in order to be able to perform a more in-depth analysis of the actual information obtained from the information modules and how function $\mathcal{F}$ works internally to build arguments and counterarguments; developments from works such as Toni (2012) and Heras et al. (2010) can be very useful in such analyses.

A very important direction of research is the extraction of preferences within, and among, information modules. This will require the analysis of how to adequately combine user behavior, the networks' structure, and methods to extract properties from the social networks such as strength of relationships, influence of nodes within and across networks, and information diffusion patterns and models. Finally, also related to preferences, we could take advantage of the rich representation capabilities of MTAFs in order to capture semantic partitions within information modules. For instance, consider an information module representing the user's network in Facebook; the module could be partitioned for instance in *Friends from Facebook*, *Activities/ Event in Facebook*, etc. This partition would allow a finer management of preferences within and across modules. For instance, if a user spends more time on Facebook than in G+, then the preferences can reflect this and give priority to information coming from the Facebook module. However, if the module *Friends from G+* is, for some reason, more important to the user, then the partition would allow to express a higher preference for information coming from that module over any other coming for Facebook.

Acknowledgments

The first author was partially supported by the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) grant EP/J008346/1 "PrOQAW: Probabilistic Ontological Query Answering on the Web" and by a Google Research Award. The second author was partially funded by the Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (Bahía Blanca, Argentina) and the Consejo de Investigaciones Científicas y Técnicas (CONICET, Argentina).

Cross-References

- ▶ Classical Algorithms for Social Network Analysis: Future and Current Trends
- ▶ Collective Classification
- ▶ Collective Classification, Structural Features
- ▶ Data Mining
- ▶ Distance and Similarity Measures
- ▶ E-Commerce and Internet Business
- ▶ Facebook's Challenge to the Collection Limitation Principle
- ▶ Flickr and Twitter Data Analysis
- ▶ Friends Recommendations in Dynamic Social Networks
- ▶ Recommender Systems, Semantic-Based
- ▶ Semantic Social Networks
- ▶ Semantic Social Networks Analysis
- ▶ Social Recommendation in Dynamic Networks

References

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749

Amgoud L, Cayrol C (2002) Inferring from inconsistency in preference-based argumentation frameworks. J Auton Reason 29(2):125–169

Baroni P, Giacomin M (2007) On principle-based evaluation of extension-based argumentation semantics. Artif Intell 171:675–700

Bench-Capon TJM, Dunne PE (2007) Argumentation in artificial intelligence. Artif Intell 171(10–15): 619–641

Besnard P, Hunter A (2001) A logic-based theory of deductive arguments. Artif Intell 128(1–2):203–235

Borgatti S, Everett M (2006) A graph-theoretic perspective on centrality. Soc Netw 28(4):466–484

Breslin JG, Harth A, Bojars U, Decker S (2005) Towards semantically-interlinked online communities. In: 2nd European conference on the semantic web: research and applications, ESWC, Heraklion. Springer, Berlin/Heidelberg, pp 500–514

Breslin JG, Passant A, Decker S (2009) The social semantic web, 1st edn. Springer, Heidelberg/New York

Bröcheler M, Pugliese A, Subrahmanian VS (2009) DOGMA: a disk-oriented graph matching algorithm for RDF databases. In: Proceedings of the international semantic web conference, Chantilly, pp 97–113

Bröcheler M, Shakarian P, Subrahmanian VS (2010) A scalable framework for modeling competitive diffusion in social networks. In: Proceedings of SocialCom/PASSAT, Minneapolis, pp 295–302

Bröcheler M, Pugliese A, Subrahmanian VS (2011) A budget-based algorithm for efficient subgraph matching on huge networks. In: ICDE workshops, Hannover, pp 94–99

Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the flickr social network. In: Proceedings of the 18th international conference on world wide web, WWW, Madrid. ACM, New York, pp 721–730

Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif Intell 77(2):321–358

Fazzinga B, Gianforme G, Gottlob G, Lukasiewicz T (2010) Semantic web search based on ontological conjunctive queries. In: Link S, Prade H (eds) Foundations of information and knowledge systems. Volume 5956 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 153–172

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41

Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1:215

FriendFeed (2013) http://friendfeed.com/

García AJ, Simari GR (2004) Defeasible logic programming: an argumentative approach. Theory Pract Log Program (TPLP) 4:95–138

Gottifredi S, García AJ, Simari GR (2011) Argument types and typed argumentation frameworks. In: Sanjay M, Nir O, Francesca T (eds) Theorie and applications of formal argumentation - first international workshop. TAFA 2011, Springer 2012 Lecture Notes in Computer Science ISBN 978-3-642-29183-8

Harth A, Umbrich J, Hogan A, Decker S (2007) YARS2: a federated repository for querying graph structured data from the web. In: Proceedings of ISWC/ASWC, Busan, pp 211–224

Heras S, Atkinson K, Botti V, Grasso F, Julian V, McBurney P (2010) How argumentation can enhance dialogues in social networks. In: 3rd international conference on computational models of argument, Desenzano del Garda. IOS, Amsterdam, pp 267–274

Jackson MO, Yariv L (2005) Diffusion on social networks. Working papers 1251, California Institute of Technology, Division of the Humanities and Social Sciences

Kang C, Molinaro C, Kraus S, Shavitt Y, Subrahmanian VS (2012) Diffusion centrality in social networks. In: International conference on advances in social networks analysis and mining, Istanbul

Mugnier M (2011) Ontological query answering with existential rules. In: 25th international workshop on description logics, Barcelona, pp 2–23

Nepomuk social semantic desktop (2008). http://nepomuk.semanticdesktop.org/nepomuk/

Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the 7th international conference on language resources and evaluation, Valletta

Peis E, Morales del Castillo JM, Delgado-López JA (2008) Semantic recommender systems. Analysis of the state of the topic. Hipertext.net 6 (online)

Prakken H (2005) Coherence and flexibility in dialogue games for argumentation. J Log Comput 15(6): 1009–1040

Prakken H (2010) An abstract framework for argumentation with structured arguments. Argum Comput 1(1):93–124

Rahwan I, Simari GR (2009) Argumentation in artificial intelligence, 1st edn. Springer, Dordrecht/New York

Ronen R, Shmueli O (2009) Evaluating very large datalog queries on social networks. In: Proceedings of the 12th international conference on extending database technology: advances in database technology, EDBT, Saint-Petersburg. ACM, New York, pp 577–587

Subrahmanian VS, Reforgiato Recupero D (2008) AVA: adjective-verb-adverb combinations for sentiment analysis. IEEE Intell Syst 23(4):43–50

The Friend of a Friend (FOAF) project (2013) http://www.foaf-project.org/

The knowledge graph (2013) http://www.google.com/insidesearch/features/search/knowledge.html

Toni F (2012) Reasoning on the web with assumption-based argumentation. In: Eiter T, Krennwallner T (eds) Reasoning web. Semantic technologies for advanced query answering. Volume 7487 of lecture notes in computer science. Springer, Berlin/Heidelberg, pp 370–386

Zhou Y, Cuenca Grau B, Horrocks I (2012) Efficient upper bound computation of query answers in expressive description logics. In: Proceedings of description logics, Rome

Query Auto-completion

▶ Weblog Analysis

Query Expansion

▶ Weblog Analysis

Query Log Analysis

▶ Weblog Analysis

Query Segmentation

▶ Weblog Analysis

Query Sessions

▶ Weblog Analysis

Query Spelling Correction

▶ Weblog Analysis

Query Suggestion

▶ Weblog Analysis

Querying and Sampling Social Networks

▶ Transforming and Integrating Social Networks and Social Media Data

Querying Volatile and Dynamic Networks

Sarvenaz Choobdar[1], Pedro Ribeiro[1], and
Fernando Silva[2]
[1]CRACS and INESC-TEC, Faculdade de
Ciencias, Universidade do Porto, Porto, Portugal
[2]Computer Science Department, CRACS and
INESC-TEC, Faculdade de Ciencias,
Universidade do Porto, Porto, Portugal

Synonyms

Clustering; Event detection; Network dynamics; Structural roles

Glossary

Dynamic (or Volatile) Network A network whose structure changes over time, some nodes and edges may appear and disappear

Local Property A metric that describes a topological aspect on the local neighborhood of a node

Structural Role The structural position of a node in the network, characterized by its local properties

Transition Pattern A typical role change, characterized by its time interval and by the origin and destination role

Definition

Many networks are intrinsically dynamic and change over time. These networks can be very volatile, with a significant number of edges and nodes appearing and disappearing. The majority of the existing network mining methodologies are however geared towards a more static scenario, with a single graph describing the topology of the system being analyzed. There is still a need for measurements and tools that allow the temporal dimension on network analysis to be fully and naturally incorporated.

In this entry we deal precisely with volatile and dynamic networks and how to extract information from them. We take a general approach in which we track the structural role of nodes over time. By using a set of local properties, such as degree and clustering coefficients, we group similar nodes and we monitor the behavior of these groups over time, looking for significant changes. We are able to describe the evolution of the network in the form of transition patterns that explain how position of nodes change and how groups of nodes grow and shrink.

Introduction

Networks are a powerful and flexible model to represent a multitude of systems. Research on network analysis, in general, and social networks, in particular, has been very prolific. There are a large number of measurements available for a practitioner wanting to understand a particular network. For a good overview see Costa et al. (2007).

The methodologies used have a broad scope and span many different conceptual metrics. Methods such as community detection (Fortunato 2010) tackle potentially large groups of nodes. Others concentrate on smaller groups of nodes, such as network motif discovery (Milo et al. 2002) or frequent subgraph mining (Huan et al. 2003). Finally, there are methodologies that try to analyze at the individual level, for instance, by classifying nodes (Macskassy and Provost 2007).

Here we focus on this last angle of approach and analyze the network from the point of view of individual nodes by measuring local properties of nodes such as clustering coefficient (Newman 2001) or the degree distribution (Dorogovtsev and Mendes 2005). Our main focus is on the temporal dimension of networks to discover how structural characteristics of nodes change over time. We use a set of local properties that fully characterize the topology of single nodes to define structural positions of nodes, and we track how these positions evolve in networks' life span. In order to better understand the global network dynamics, we group the nodes based on structural properties and we observe the temporal behavior of nodes. In this entry we are trying to answer questions such as the following: What structural positions or roles do exist in a network? How are they appearing and disappearing? Are specific roles growing or shrinking? Which periods of time show significant changes and which groups of nodes exemplify such changes? Moreover, we are also interested in understanding the transitions between roles of nodes. Which role is losing nodes to another? What is the typical path taken by the structural role of nodes? Can we pinpoint in time and identify major role transitions?

Key Points

Our proposed methodology is geared towards a dynamic network that keeps a relatively static set of nodes (so that we can track them over time), but has a volatile set of edges, that keep changing and evolving. We are able to characterize single nodes using local properties, to group them by structural function using clustering over the properties, and to discover significant events, positioning them in time by discovering the relevant time intervals and identifying the relevant transition patterns by creating sets of rules that describe the change of nodes from one role to another.

Historical Background

Only recently, more attention is being given to the dynamic aspects of networks. From a global perspective, research has been devoted to the study of general growth patterns of networks (Leskovec et al. 2005, 2008; Sun et al. 2006) and, on a less global scale, to aspects such as the formation and evolution of communities (Greene et al. 2010; Lin et al. 2008; Tang et al. 2008), or subgraph dynamics (Berlingerio et al. 2009; Jin et al. 2007). On a more local level, link prediction has also been used to guess appearance of edges over time (Liben-Nowell and Kleinberg 2007).

Modeling the temporal behavior of nodes is a relatively emergent point of view. In this line of research, the main focus is on the single nodes and their properties in the network. In our previous works (Choobdar et al. 2011, 2012) a two-phase general methodology was designed to characterize time-evolving networks. In the first step of this methodology, nodes are grouped by k-means clustering and classified based on their role in the network. In the second step a method is proposed to study the evolution of the network by a supervised approach. In this method a set of events happening in the network is defined for the roles in the network. We then find the predefined events happening in the network and the rules that describe them by using association rule mining. On a related approach, Rossi et al. (2012) used the methodology proposed by Henderson et al. (2012) for static role extraction. They measure a set of features for nodes at each time snapshot; then by stacking all the node-by-feature matrices, they derive the matrix of feature roles by factorizing the stacked node-by-feature matrix and iteratively generate the matrix of node role for each time.

Proposed Methodology

In this section we formulate the problem of dynamic network study from a single node point of view. Our goal is to model the temporal characteristics of nodes regarding the different structural positions they hold, which we call *roles* in this entry. For a given sequence of network snapshots, we examine the dynamics by tracking roles of nodes over time. For a given node, the role is defined regarding properties in its close neighborhood. The proposed method for this problem consists of three steps: (1) measuring local properties of nodes at each time step, (2) extracting structural roles of nodes, and (3) finding patterns of role transition.

Measuring Local Properties

The role of a node is automatically determined based on its properties in the network. Nodes are assigned to the same role if they are similar. We assess the similarity of two nodes by using their properties rather than number of edges between them. Two nodes are close if they have a similar *feature vector*. Initially, we need to select a set of local measurements that best characterize nodes in the network structure. We chose to employ the same set of metrics used by Costa et al. (2009) as the feature vector for finding groups of nodes. These features measure the connectivity of a node in the neighborhood structure. In particular, we use the following five properties:

- The normalized average degree (r)
- The coefficient variation of the degrees of the immediate neighbors of a node (cv)
- The clustering coefficient (cc)
- The locality index (loc), which is an extension of the matching index and takes into account all the immediate neighbors of each node, instead of individual edges
- The normalized node degree (K)

Extracting Structural Roles

To extract the set of roles describing the existing positions in the networks, we use the well-known k-means clustering algorithm (Hartigan and Wong 1979) to find groups of nodes with similar feature vectors. Each cluster contains nodes with a similar position in the network regarding their feature vectors, hence the same role or label can be assigned to them.

Given a sequence of graphs $\{G_1, G_2, \dots, G_t\}$, we build the feature matrix M_t by calculating the local properties of nodes at time t. We discover groups of nodes by applying k-means algorithm on the stacked matrix of all node-feature matrices up to time t. In the stacked matrix, one node may appear multiple times with different feature vectors, if it is active (i.e., if it exists) at multiple time steps. We derive a membership matrix (node by time) for all nodes of the network regarding the clustering result that shows the role of nodes at each time step.

The number of potential groups of nodes in the network is equal to the number of clusters in the dataset. Determining the actual number of groups in a dataset is a fundamental and largely unsolved problem in cluster analysis. We employ the method by Sugar and James (2003), since it does not require parametric assumptions, is independent of the method of clustering, and was shown to achieve excellent results. This method uses a theoretic information approach that considers the transformed distortion curve $d : K^{-p/2}$, where p is the number of dimensions in the dataset.

Role Transition Modeling

We explore the dynamic of a network at node level by finding role transition patterns. The temporal behavior is modeled to see if nodes hold similar roles over their lifetime or consistently oscillate over time. For example, in a social network, a center of a star was a bridge node before or is it likely to turn to an isolated node in its life span? We generate a set of transition rules to describe these events in a network.

Five basic types of temporal patterns are defined for a role regarding the number of nodes having that role: (1) growth, (2) shrink, (3) emergence, (4) dissolution, and (5) constant. We learn these patterns by discovering transition intervals where the number of nodes of a certain role is shrinking, growing, or showing other defined temporal patterns. After this, a set of rules describing the pattern is generated.

Transition Intervals

A transition interval is a time interval where a considerable number of nodes leave or join a role. For a given role C_i, its size over time constitutes a time series denoted as $F_i(t), t \in [1, T]$. A transition interval is the subsequence of F_i which holds a constant increasing or decreasing trend. Hence, we extract transition intervals of F_i by segmentation of the time series. Starting from $t = 1$, $F_i(t)$ is approximated by linear regression to find the transition intervals. If the error of the fitted line for a subsequence $F_i[a : b]$ exceeds the threshold, the interval $[a, b]$ breaks to point j where it gives the best approximation for $F_i[a, j - 1], j < b$. The error is measured in terms of the sum square of residuals. The threshold is controlled by the maximum number of arbitrary transition intervals. The maximum error is increased while the number of intervals is below the defined maximum number of intervals. For this method we can define either the maximum error for the linear regression or the

Querying Volatile and Dynamic Networks, Table 1 Dataset statistics: number of time snapshots, number of nodes and edges at the final snapshot, and node and edge growth rate (ratio between the final and initial time snapshots)

Dataset	Time snapshots	$\|V\|$	$\|E\|$	Node growth rate	Edge growth rate
Global trade	53	186	8,839	2.47	7.93
USA airports	244	1,919	14,391	1.64	1.21
DBLP	11	31,592	49,599	3.4	4.57

maximum number of desired intervals. The slope of the fitted line for each segment shows whether the interval is increasing or decreasing, which respectively determines the growth (emergence) or shrink (dissolution) events.

Transition Rules

We extract a set of rules to describe the specific pattern at transition intervals. A transition rule is in the form of $C_i \rightarrow C_j$, which shows that nodes from group i moved to j. The order of transitions is important, since it shows the trend of changes in the network properties. For example, for a sequence of $\{1, 1, 1, 2, 2, 3\}$, the transition rules are $1 \rightarrow 2$ and $2 \rightarrow 3$. We extract these one-step transition rules for a time interval by building a transition matrix in each time interval. The support count of a rule $C_i \rightarrow C_j$ is defined as the number of nodes that go from cluster C_i to C_j in that interval.

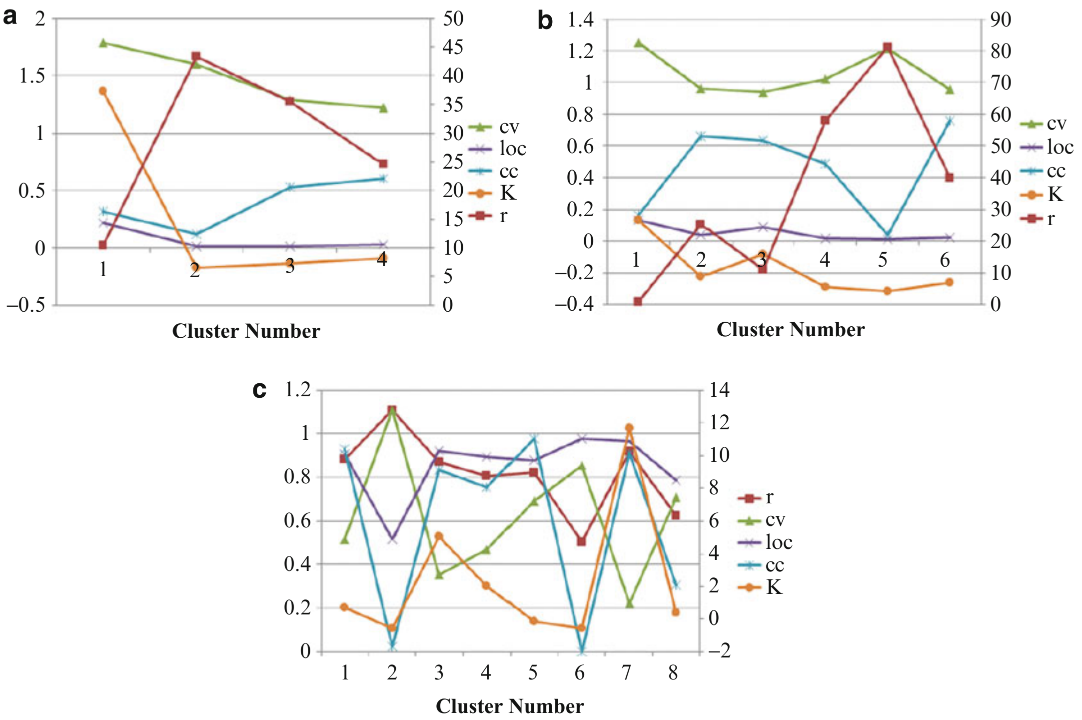
Exploratory Analysis

We now apply our proposed methodology to real data in order to show its applicability. For our experiments, we used three different real complex networks: world countries global trade (Gleditsch 2002) (two countries A and B are connected at time t if the trade between them is more than 10 % of either country's total trade in that year), connections between USA airports (Route Views 1997), and a coauthorship network obtained from DBLP data (Berlingerio et al. 2009). We use the undirected forms of these networks. Table 1 gives some more details on the topology of three used networks.

This methodology gives a profile for each network, including information about the existing roles and the temporal patterns, describing the existing dynamics. Figure 1 shows the profile of the roles found in each network. The profile depicts the values of the feature vector of each role in the network. This figure clearly depicts that each role in the network has a different feature vector and that nodes are distinguished according to their structural position.

For example, in the global trade network of countries, our method found four distinguishable roles for nodes. Nodes with role number 1 are countries with very high-degree and many low-degree nodes connected to them. Neighbors of these nodes have low degree since the normalized average degree of the immediate neighbors of a node for this group is very low. This means that nodes in role 1 behave as hubs in the network, i.e., as hub countries in global trade, with commercial transactions with many other countries that have a high variation of degree in neighborhood (cv). According to the value of (loc) and (cc), respectively, the locality index and the clustering coefficient, nodes are highly connected in their neighborhood. The United States of America, Canada, and France are members of this group.

Table 2 shows some of the extracted patterns. Besides the frequency count of every pattern in each dataset, and in order to assert their significance, we compared the result with randomized sequences. We built these random sequences by shuffling the order of role membership of each node in the network. In this way, the random dataset has the same number of nodes and role types. Since we have a dataset of role sequences for each network, we built 10 datasets of random sequences with the size as the number of nodes. We calculated the average and standard deviation of frequencies for each rule of events in the 10 datasets. Finally, we computed the z-score for the significance of each rule as



Querying Volatile and Dynamic Networks, Fig. 1 Illustration of the feature vector of the roles in the selected networks. The normalized average degree (r) is plotted on the second axes on the right side of the plot (**a**) Global trade network (**b**) USA airports network (**c**) DBLP network

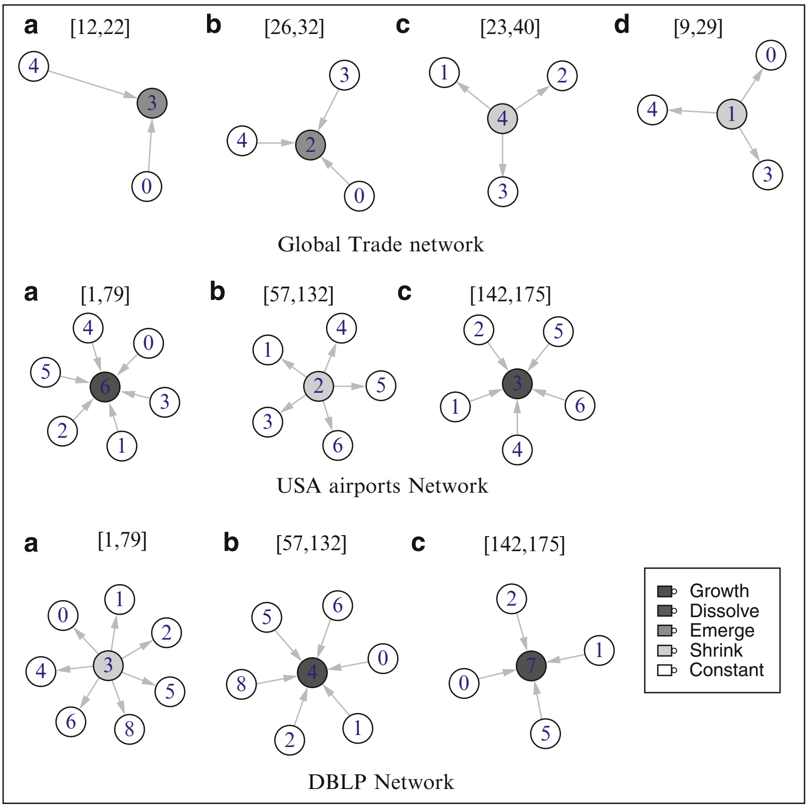
Querying Volatile and Dynamic Networks, Table 2 Description of some extracted events in the networks. Numbers in the parenthesis denote the cluster number, holding the events

Dataset	Event	Time interval	Transition rules	Support	Z score
Global trade	Emerge (2)	[26,32]	3 → 2	80	6.3
	Shrink (4)	[23,40]	4 → 3	100	14.70
	Emerge (3)	[12,22]	4 → 3	126	8.3
	Shrink (1)	[9,29]	1 → 4	25	12.1
USA airports	Growth (3)	[142,175]	1 → 3	1,062	3.75
	Growth (6)	[1,79]	2 → 6	437	7.27
	Shrink (2)	[57,132]	2 → 3	551	4.79
DBLP	Growth (4)	[1,11]	0 → 4	6,710	1.2
	Growth (7)	[6,11]	0 → 7	196	0.2
	Shrink (3)	[1,7]	3 → 0	726	1.3

compared to the randomized form. The table shows the rules with the highest z-score.

Figure 2 graphically illustrates some of the role transition patterns happening. Each graph includes all role transitions in a specified time interval. The label of nodes specifies the role and the color of nodes shows the type of temporal pattern.

For example, the first event (a) in the global trade network describes the emergence of cluster 3 in the [12,22] time interval. This graph says that nodes that constitute cluster 3 either come from cluster 4 or are the new nodes (cluster 0), just joining the network. By observing Table 2, we can see that the main reason for this event is the transition from cluster 4.



Querying Volatile and Dynamic Networks, Fig. 2 Some of the events happening in the networks

The same type of interpretation could be applied for the two other networks.

Regarding the scalability of our approach, our methodology gives results in a reasonable time for small-scale networks like global trade or a relatively large network like DBLP. The size of the network essentially affects the first step of the proposed methodology where the local properties of nodes are calculated. In this entry, this set of properties can be computed efficiently since they are based on one 1st degree ego network of nodes. The other steps of the method utilize standard machine learning techniques. The clustering step uses k-means, which is an NP-hard problem, but there are efficient approximate solutions (Jain 2010). The role transition modeling step is more affected by the number of time snapshots rather than the size of network since transition rules are derived by matrix operations in every time step.

Key Applications

This method provides a general picture of a network and its dynamics over time and as such has broad applicability. The results of our methodology are helpful for comparing networks regarding their role distribution. For example, we can examine if there is the same role distribution in social and biological networks. Finding outliers in a network is another possible application. Nodes that exhibit a temporal behavior different from the other nodes in their discovered role set can be considered as outliers.

Future Directions

In the near future we intend to study the usage of other types of measurements. For instance, can subgraph metrics, such as participation of nodes in network motifs, be used for role discovery? We also intend to study the applicability of our methodology to very large-scale networks, by scaling up our used algorithms. In addition, we plan to apply the proposed method for network studies such as network comparison, node classification, and anomaly detection.

Acknowledgments

This work is in part funded by the ERDF/COMPETE Programme and by FCT within project FCOMP-01-0124-FEDER-022701. Sarvenaz Choobdar is funded by an FCT Research Grant (SFRH/BD/72697/2010). Pedro Ribeiro is funded by an FCT Research Grant (SFRH/BPD/81695/2011).

Cross-References

- ▶ Analysis and Visualization of Dynamic Networks
- ▶ Community Identification in Dynamic and Complex Networks
- ▶ Path-Based and Whole-Network Measures
- ▶ Role Discovery
- ▶ Role Identification of Social Networkers
- ▶ Temporal Analysis on Static and Dynamic Social Networks Topologies

References

Berlingerio M, Bonchi F, Bringmann B, Gionis A (2009) Mining graph evolution rules. In: Machine learning and knowledge discovery in databases, Proceedings of ECML PKDD 2009, Bled Slovenia, Springer, LNCS 5781, pp 115–130

Choobdar S, Silva F, Ribeiro P (2011) Network node label acquisition and tracking. In: Progress in artificial intelligence, 15th portuguese conference on artificial intelligence – EPIA'11, Lisbon. Lecture notes in artificial intelligence, vol 7026. Springer, Heidelberg, pp 418–430

Choobdar S, Silva F, Ribeiro P (2012) Event detection in evolving networks. In: Fourth International Conference on Computational Aspects of Social Networks, CASoN 2012, IEEE, Sao Carlos, Brazil, pp 26–32 doi: 10.1109/CASoN.2012.6412373.

Costa LF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: a survey of measurements. Adv Phys 56:167

Costa L, Rodrigues F, Hilgetag C, Kaiser M (2009) Beyond the average: detecting global singular nodes from local features in complex networks. Europhys Lett 87:18008

Dorogovtsev SN, Mendes JFF (2005) The shortest path to complex networks. In: Complex Systems and Inter-disciplinary Science', v. 1, eds. N. Johnson, J. Efstathiou, and F. Reed-Tsochas

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Gleditsch K (2002) Expanded trade and GDP data. J Conf Resolut 46(5):712

Greene D, Doyle D, Cunningham P (2010) Tracking the evolution of communities in dynamic social networks. In: International conference on advances in social networks analysis and mining (ASONAM), Odense. IEEE, Odense, Denmark, pp 176–183

Hartigan J, Wong M (1979) A k-means clustering algorithm. J R Stat Soc C 28(1):100–108

Henderson K, Gallagher B, Eliassi-Rad T, Tong H, Basu S, Akoglu L, Koutra D, Faloutsos C, Li L, Matsubara Y et al (2012) Rolx: structural role extraction and mining in large graphs. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining, ACM, Beijing, China, pp 1231–1239

Huan J, Wang W, Prins J (2003) Efficient mining of frequent subgraphs in the presence of isomorphism. In: Proceedings of the third IEEE international conference on data mining, ICDM'03, Melbourne, Florida, pp 549–552

Jain AK (2010) Data clustering: 50 years beyond k-means. Pattern Recognit Lett 31(8):651–666

Jin R, McCallen S, Almaas E (2007) Trend motif: a graph mining approach for analysis of dynamic complex networks. In: 7th IEEE international conference on data mining, Nebraska. IEEE, Los Alamitos, pp 541–546

Leskovec J, Kleinberg J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the 11th ACM SIGKDD international conference on knowledge discovery in data mining, Chicago, USA, pp 177–187

Leskovec J, Backstrom L, Kumar R, Tomkins A (2008) Microscopic evolution of social networks. In: Proceeding of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, Las Vegas, pp 462–470

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol (JASIST), Wiley, 58(7):1019–1031

Lin Y, Chi Y, Zhu S, Sundaram H, Tseng B (2008) Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: Proceedings of the 17th international conference on World Wide Web, Beijing. ACM, Beijing, China, pp 685–694
Macskassy S, Provost F (2007) Classification in networked data: a toolkit and a univariate case study. J Mach Learn Res 8:935–983
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594): 824–827
Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E, 64(1):016131
Rossi R, Gallagher B, Neville J, Henderson K (2012) Role-dynamics: fast mining of large dynamic networks. In: Proceedings of the 21st international conference companion on World Wide Web, ACM, Lyon, France, pp 997–1006
Route Views (1997) University of Oregon Route Views Project. Online data and reports. http://www.routeviews.org. Accessed Feb 2013
Sugar C, James G (2003) Finding the number of clusters in a dataset. J Am Stat Assoc 98(463):750–763
Sun J, Tao D, Faloutsos C (2006) Beyond streams and graphs: dynamic tensor analysis. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, ACM, Philadelphia, USA, pp 374–383
Tang L, Liu H, Zhang J, Nazeri Z (2008) Community evolution in dynamic multi-mode networks. In: Proceeding of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, Las Vegas. ACM, Las Vegas, USA, pp 677–685

Questionnaires for Measuring Social Network Contacts

Tina Kogovšek[1], Valentina Hlebec[2],
Germà Coenders[3], and Lluís Coromina[3]
[1]Faculty of Arts and Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia
[2]Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia
[3]Department of Economics, University of Girona, Girona, Spain

Synonyms

Data collection methods; Quantitative measurement; Surveys

Glossary

Interaction Approach Data collection approach where respondents are required to record all social contacts (interactions) in a designated time frame

Role-Relation(ship) Approach Data collection approach where the criterion for generating a social network is social roles

Affective Approach Data collection approach where the criterion for generating a social network is the subjective value of ties for the respondents

Exchange Approach Data collection approach where the criterion for generating a social network is exchanges of resources

Network Generator A question for generating a social network

Name Generator A type of a network generator which elicits names of actual persons

Role Generator A type of a network generator which elicits persons as social roles

Resource Generator A type of a network generator which elicits persons who are in specific positions in a society or possess specific skills or resources

Name Interpreter A question for measuring properties of network members or properties of ties between the respondent and a network member

Boundary Specification Problem A question of criteria for inclusion of persons in a social network, usually related to whole networks

Free Recall Method A method of eliciting network members free from memory

Recognition Method A method of eliciting network members from a list of persons provided in advance

Definition

In social network analysis, there are several very different data sources, obtained by various data collection methods – qualitative and quantitative:

- Survey
- Archive

- Participant observation
- Informants
- Diary
- Electronic sources
- Other sources

The most frequently used method is survey, which is also the focus of this entry. Surveys use structured questionnaires to gather information directly from the network members and thus belong to the domain of empirical quantitative research methods.

Different survey data collection methods are used for collecting network data, most often face to face, but also telephone and mail and recently increasingly web data collection. Network data collected by surveys have the same advantages and disadvantages as "traditional" survey data (e.g., opinions, attitudes) but additionally face some additional challenges owing to the specifics of collecting data on relations among people. Some of these are described in this entry. Additionally, the interested reader is referred also to the entry on quality of network data and the entry on collection of network data on the web.

Introduction

Information on relationships among people, institutions, and other types of actors in a social network can be gathered with a wide array of methods.

Archival data, usually different sorts of historical data, can be a valuable source. In the field of social network analysis, an example could be data on relations among the Florentine families (e.g., Padgett and Ansell 1993).

Participant observation, informants, and, to some extent, diaries are most frequently used in qualitative studies. Often these data collection methods are subsumed under the term fieldwork, where the researcher is himself/herself part of the research setting, staying among the research participants for a prolonged period of time and usually using a combination of the mentioned methods and sometimes even participating in as one more network member (participant observation). One of the well-known examples of qualitative network data are the Sampson monastery data (Sampson 1968).

Electronic data have become more often used after the rapid spread of the Internet and include data on exchanges of messages by email and data on relations among people on web-based social networks, such as Facebook and similar.

Other sources can even include nonnetwork kinds of data, out of which network data can be generated. Examples could be networks of words in a text or a dictionary (e.g., in linguistics) or networks of web pages.

This notwithstanding, surveys using structured questionnaires are a commonly used systematic and structured way to gather network data from the responses of network members themselves, which may be used as a stand-alone research method or be combined with the abovementioned alternative data sources.

Key Points

We could start discussing survey network data collection by describing the four most general approaches to measuring social networks (e.g., Van der Poel 1993). They represent the most general methodological framework for eliciting the networks, i.e., what kind of a stimulus is used in a survey question to obtain the network. Social networks can be defined in the sense of simple interactions, i.e., simply with whom respondents have had contacts over some period of time, that would be the interaction approach. Another way of thinking about social networks would be the formal, normative structure of a society, consisting of roles which we perform in different life situations, such as being a parent, being an employee, being a customer, or being a supplier; this is then the role-relation(ship) approach. Thirdly, social networks can be organized around the concept of emotional attachment – who are the people we feel close to, who are important to us, and who we perceive as more or less distant – this is the affective approach. And finally, we can think of social networks as a means of exchanges of different kinds of resources, material or nonmaterial, such as talking about personal problems, helping with household tasks, socializing, and getting industry know-how, political connections, or financial resources. This is then the exchange approach.

Empirically, these approaches are then operationalized in different types of actual survey questions for obtaining respondents' social networks – the network generators. These networks can be elicited as names of actual persons (e.g., John F., Maria L.), social roles (e.g., mother, partner, friend), social positions (e.g., owner, manager), or persons possessing specific resources (e.g., specific skills, knowledge). Additionally, questions about characteristics of these ties and network members themselves can be obtained.

In surveys, ego-centered social networks are usually studied, i.e., ties of respondents (usually called egos) to their network members (usually called alters). Therefore, the examples in this entry are also centered around this type of networks but can be easily applied to whole networks (i.e., a more or less finite group of people with a set of relations defined among them such as a class or a company). In whole network analysis, members can be asked about their own networks in a similar manner as in ego-centered networks, the whole network being later compounded with all reported ties around each actor. It can also be the case that members are asked to report relationships among third parties in the whole network, which is known as proxy reporting.

Historical Background

The beginnings of social network data collection and analysis can be traced as far back as the 1930s when sociometry appeared within the field of social psychology, founded by Jacob Moreno. Moreno was especially interested in group dynamics. He developed the so-called sociogram as a method of analyzing and understanding the structure of relations in a group and which was the first (paper-and-pencil) questionnaire for collecting network data (Moreno 1934). A sociogram is a graphical representation of a social network, where persons are represented as points (nodes) and relations as lines among them. Within this framework most formal analytical procedures for social network analysis were developed (e.g., Wasserman and Faust 1994).

Historically, most questionnaires for social network data were paper and pencil and administered in face-to-face interviews. More recently, in the last decade or so, also other types of data collection are being used to a larger degree such as telephone surveys and, increasingly, web surveys. As in nonnetwork type of studies, with the development and spread of ever more powerful computers, computer-assisted questionnaires have become standard also in collecting network data. Owing to inherent complexity of collecting network data, computer-assisted questionnaires are especially valuable and useful.

In the present time questionnaires for obtaining social networks data are used in very different fields of research, such as epidemiology, organizational research, political science, sociology, and economics. A field of research where social network concepts have a very important part and where data are mainly collected by surveys is the field of social support (for a good overview of basic concepts, see Vaux 1988). Therefore, also most of the examples in this entry are taken from this field but can be easily applied also to other social contexts, such as organizational research (Coenders and Doreian 2011).

Data Collection Approaches

Let us start by the fact that survey questionnaires for measuring social networks are in some respects similar to the "traditional" questionnaires. In social network questionnaires, more general social science concepts, such as attitudes, opinions, and values, are often measured as well. Usually this is done by way of Likert-type scales which are formulated as statements which respondents evaluate in closed format (e.g., five-point scale of (dis)agreement). On the other hand, measuring social networks has at least one important specific characteristic, namely, we are measuring social relations which require special types of questions for eliciting these relations and which are presented later on in this entry.

Let us now proceed by describing the four most often used approaches to network data collection (e.g., Van der Poel 1993). Each one of them consists of a specific variety of a stimulus or criterion for inclusion of network members into a social network. These approaches are interaction, role-relation, affective, and exchange approaches.

Within the *interaction approach* respondents are usually asked to keep a record of all social contacts for a certain period of time (e.g., Milardo 1989; Van der Poel 1993). In this way, all people with whom a respondent has had social interactions are identified. This approach is relatively simple. But its main disadvantage is its lack of specific context within which an interaction takes place. For instance, by this approach a lot of substantively relatively unimportant contacts may be obtained (e.g., contacts with the postman, neighbors, mere acquaintances).

The *role-relation approach* uses role relationships as a criterion for generating a social network (e.g., Van der Poel 1993). Role relationships (or social roles) are a well-known and often used concept in sociology and represent specific sets of expectations, obligations, and rights into which individuals enter throughout their lifetimes in different situations, such as having children and obtaining the role of father/mother or getting a job and becoming an employee and similar. Role relationships are therefore used to describe normative or formal aspects of relationships. The main advantage of the approach is in its everyday familiarity for respondents; however, it tends to give more importance to family roles and to underestimate other types of relations such as coworkers, fellow members of associations, and acquaintances.

The *affective approach* uses the subjective value which a relationship has for the respondent, such as closeness, intimacy, or importance as the criterion for inclusion into the network (e.g., Van der Poel 1993). The affective approach was first proposed by Kahn and Antonucci (1980) where the criterion of closeness is used to distinguish three circles of network members radiating outwards from the respondent who is in the center. Along the dimension of closeness, characteristics of network members change with respect to closeness, composition (percentages of family, friends, etc.), and duration. The biggest advantage of this approach is at the same time its biggest disadvantage – the respondent himself/herself defines the meaning of closeness. On the one hand, the researcher avoids enforcing his/her definitions of closeness that may not have particular relevance for the respondent. On the other hand, respondents may use different criteria/definitions of closeness either individually or collectively (e.g., in different cultures) to which the researcher has no access and may compromise comparability of data. For instance, "an important person" may mean "family" to one respondent, "someone with whom I have frequent contact with" to another, and "someone to whom I am emotionally close" to another.

With the *exchange approach* first proposed by McCallister and Fischer (1978), network members with whom respondents exchange different types of resources are obtained. These resources may be emotional (e.g., discussing personal problems, providing encouragement, building self-esteem), instrumental (e.g., helping in household, borrowing money, taking care of children), informational (providing information during important life changes such as looking for a job, searching for a new apartment), or social companionship (e.g., doing sports together, visiting, going to the cinema). The approach provides a substantively richer, more detailed, and broader picture of the respondent's social network and reduces the possibility of subjective definitions of important relationships but is on the other hand more complex and therefore potentially more demanding to operationalize and administer.

Several studies (e.g., Milardo 1989) compared and discussed these four types of approaches to obtaining social networks. It seems, when comparing the affective and the interaction approaches, the social networks yielded by these two approaches differ substantially. The affective networks were much smaller than the interactive networks. There was only a 25 % overlap in network membership. The affective approach seems to elicit relatively small, strong-tie, affective networks. In general cited network members (data were collected by face-to-face interviews) follow decreasing order with regard to closeness and frequency of contacts, except for coworkers who are contacted on a daily basis.

A comparison of the exchange, affective, and role-relationship approaches (Van Sonderen et al. 1990) has shown the exchange approach listed 78 % of all listed network members, the affective approach listed 49 %, while the role-relationship approach provided 34 % of all listed network members. In general it seems that the exchange approach elicits more network members, a larger span of different role relationships than the affective approach, and more frequently contacted network members. The affective approach tends to elicit network members whose importance is valued higher and more long-term relationships.

As is evident in this section, the selection of an appropriate approach is extremely important. Each of the described approaches captures a different range of relationships. Each of the approaches has its own advantages and disadvantages, and their use should depend on the aim of the research project the researcher is planning. When selecting the most efficient approach, one has to be certain which kinds of relationships are of primary interest.

Data collection approaches are combined with different data collection techniques which are described in the next section.

Key Applications

Data Collection Techniques

Within the field of social network analysis, several ways of measuring social networks and their characteristics by survey questionnaires are used. We call them data collection techniques and they are single and multiple name generators, role-relation generator, and resource (position) generator for generating networks (therefore, we could also call them network generators). We will review these techniques in the context of ego-centered networks, but they can be easily applied to whole networks as well.

Most often, when evaluating ego-centered social networks, one of the above network-generator methods are used. The data collection goes as follows. First, a list of respondents (egos) is obtained. In the second step, ties are identified by the network generators – all network members (alters) with whom the focal ego has some sort of relationship. Third, the contents and the characteristics of ties (e.g., closeness, importance, duration) and of the alters (e.g., gender, age, education) are usually evaluated. Let us first describe different types of network-generating questions.

One of the most often used network-generating questions is the *name generators*. They can be either single or multiple. As already the name suggests, a *single name generator* consists of only one question for obtaining the names of network members. A typical example is the well-known and often used Burt name generator (Burt 1984):

> From time to time, most people discuss important personal matters with other people. Looking back over the last six months – that would be back to last August – who are the people with whom you discussed an important personal matter?

The advantages of the single name generator are its simplicity, low cost, and low respondent burden. On the other hand, such a name generator is often defined more broadly which brings on a larger possibility of its different interpretations by individual respondents. For instance, it was shown (e.g., Bailey and Marsden 1999) that different respondents interpreted the term "important matters" in the Burt name generator differently (e.g., frequency of contact, intimacy of the relationship). A single name generator also captures only one segment of a network (e.g., the segment providing emotional support) and thus giving a limited picture of an individual's total support network. On the other hand, if our research purpose is limited and/or the social network questions are only a part of an already large survey questionnaire, using a single name generator is a valid and justifiable decision.

A multiple name generator on the other hand consists of two or more network-generating questions. An example (a shortened version taken from Kogovšek et al. (2002) and roughly based on the first such example in McCallister and Fischer (1978)) could be:

1. From time to time people borrow something from other people, for instance, a piece of equipment, or ask for help with small jobs in or around the house. Who are the people you usually ask for this kind of help?
2. From time to time people socialize with other people, for instance, they visit each other, go together on a trip or to a dinner. Who are the people with whom you usually do these things?
3. From time to time people discuss important personal matters with other people, for instance, (a) if they quarrel with someone close to them (b) when they have problems at their work or something similar. Who are the people with whom you discuss personal matters that are important to you?

Each of these questions therefore captures one segment of a network, for instance, a group of people providing a specific kind of social support to the respondent. Criteria for inclusion of alters into the network can thus be usually more clearly defined and enable identifying a much larger portion of an ego-centered network than a single name generator. Studies (e.g., Granovetter 1973) have shown that some network members, especially not the closest ones, can be very important support providers, but only in very specific situations. These network members usually turn up in a network when multiple name generators are used. A network obtained by a multiple name generator is thus usually more representative of the respondent's total network, provides richer information on the network's characteristics, and enables more valid conclusions. On the other hand, its main disadvantages are larger complexity, higher cost, and higher respondent burden. Also, the content of the questions within the multiple name generator can be very specific and therefore limiting comparability among different studies. The content of individual questions may also not be equally relevant for all respondents, possibly leading to a larger degree of missing data and errors. The questions may also be redundant by not providing any new information, but increasing respondent burden. Some research (e.g., Van der Poel 1993) also shows that using three to five name-generating questions is a relatively good compromise between obtaining as rich information about social networks as possible and economy of use.

In general the name-generator method yields data of high quality. However, it is very time and money consuming, and it requires either a considerable effort on the respondent's part or a complex coordination between interviewer and respondent when it is applied in personal interviews (e.g., Kogovšek et al. 2002; Kogovšek and Ferligoj 2004, 2005).

Data about ego-centered networks can also be collected in a simpler way, by asking an ordinary survey question where response categories are types of relationships (e.g., partner, parents, children, friends). This data collection technique can be named *the role generator* as it generates roles instead of names. An example (used, for instance, in the European Quality of Life Survey) could be:

Say you have the flu and have to rest for a few days. You need help with various household tasks such as shopping and similar.

A. Who would you ask for help first?
B. Who would you ask for help second?

	First	Second
Family member	01	01
Coworker	02	02
Friend	03	03
Neighbor	04	04
Someone else	05	05
No one	06	06

It is very appealing as it saves time and money. However, information obtained by this approach is very limited. There is no information on network size, there is no way to distinguish individual network members (such as friends Eva and Martin), there is no information regarding the quality of ties linking ego and his/her alters, and there is no information about alters' characteristics (apart from the type of relation between ego and alters). However, if one needs only limited information about support provision, such as composition of social support networks or a list of support providers and/or this is only one segment of a larger survey, a role generator may be a good decision to be used in a survey. A similar type of role generators is used in the

Questionnaires for Measuring Social Network Contacts, Table 1 An example of a resource generator

1. Do you know anyone who. . .	No	Family member	Friend	Acquaintance	
2. . . . and are you someone who. . .					yourself?
(a) . . . can repair a car, bike etc.	0	1	2	3	4
(b) . . . owns a car	0	1	2	3	4
(c) . . . is handy repairing household equipment	0	1	2	3	4
(d) . . . is active in a political party	0	1	2	3	4
(e) . . . has good contacts with a newspaper, TV- or radio station	0	1	2	3	4
(f) . . . has knowledge about financial matters (e.g., taxes, subsidies)	0	1	2	3	4

International Social Survey Programme, but a longer list of possible roles is used (especially with a more elaborate list of family roles, such as mother, father, brother, son).

One of the newer data collection techniques developed in the last decade or so in the context of studying social capital is also *the resource generator* (e.g., Van der Gaag and Snijders 2005). The resource generator measures relations of the respondent towards persons on specific societal positions or having specific skills, competences, or resources (e.g., high income, knowledge of repairing cars, expertise in a specific field of knowledge, e.g., law). Such a network generator does not provide the names of actual persons but measures respondent's access to different material or nonmaterial resources, possessed by the network members and which can be respondent's competitive advantage in different life situations. A somewhat shortened example of a resource generator (Van der Gaag and Snijders 2005) could be (Table 1):

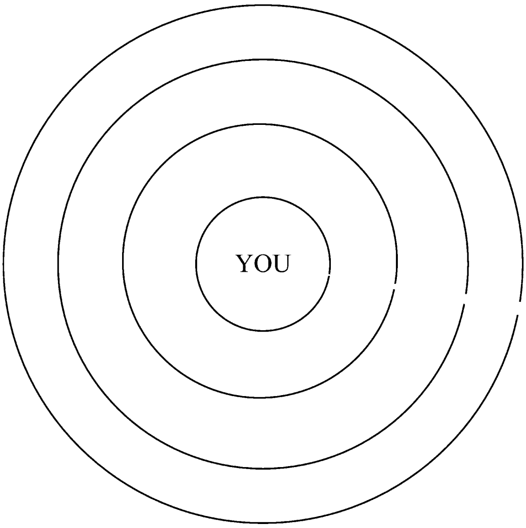
Since the resource generator does not include generating the ego-centered network, it is relatively easy and cheap to administer and does not pose a large burden on the respondent. On the other hand, availability, visibility, and appropriateness of resources tend to be very culturally specific, so one has to be very clear in research purpose and theoretical backgrounding of the planned study.

A similar concept to the resource generator is the position generator, where the respondent is asked about knowing persons in specific social positions/roles, e.g., lawyer, doctor, manager, and engineer (e.g., Lin et al. 2001).

If the researcher plans to use face-to-face or web data collection methods, then visual techniques of network elicitation and representation (e.g., Kahn and Antonucci 1980; Hogan et al. 2007; McCarty et al. 2007) can be very useful, since they tend to be intuitively appealing to the respondent and his/her inner mental representations of his/her social networks. As an example, the Kahn and Antonucci graphical representation of a network is presented in Fig. 1. The respondent's network members are elicited by three name generators, one for each circle, and the criterion for inclusion into the network is closeness/importance of a network member.

Hypothetical and Actual Network Generators

All network generators can ask about either actual or hypothetical relations. By the first names of network members actually contacted by the respondent in a specific time frame are obtained (e.g., the Burt name generator above). The advantage of such a wording is that of course actual interactions are recorded. On the other hand, such questions are plagued by the same kind of problems with recall from memory as also the



Questionnaires for Measuring Social Network Contacts, Fig. 1 The Kahn and Antonucci network

nonnetwork kind of survey questions have. Additionally, a network generator asking about actual interactions may well miss important persons that have not been available to the respondent in the designated time frame owing to specific life situations (e.g., work or study abroad and similar).

On the other hand, hypothetically worded network generators (e.g., the European Quality of Life Survey example above) have their own disadvantages too. Respondents may interpret them in the sense of naming persons that *should* be named therefore rather capturing implicit cultural norms and expectations than who they actually seek for help. Despite these potential problems, hypothetically worded network generators most probably capture the relations within which interactions actually have happened or most probably would happen and are important for respondents even if they have not had actual contact with them in a certain time frame (Van der Poel 1993). Again, the choice of the appropriate wording depends primarily on the research purpose.

Using Limitations or Not

Besides the time frame limitation, another kind of limitation, specific to social network survey questions, is limitation of the number of choices (network members). Research results in this regard are varied and inconclusive. Although some studies have shown that limitation of the number of choices causes different properties of the network (e.g., size, structure, composition) on the other hand, at least in some situations limitations do not produce largely different network properties (e.g., Holland and Leinhardt 1973; Kogovšek and Hlebec 2005; Hlebec and Kogovšek 2013). Placing no limitation has its own problems – potentially enlarging interviewer effect (different additional explanations of the meaning of the question, e.g., the definition of "friends" in the example of the resource generator above), larger complexity of the network (e.g., different kinds of "friends" in the network the researcher is not aware of), and more respondent burden owing to a larger named network.

Boundary Specification

As already mentioned, these techniques can easily be applied also to whole networks. However, two topics related more specifically to measurement of whole networks should at least be briefly mentioned. One is the so-called boundary specification problem (see Laumann et al. 1983) – who is to be included as a member of a whole network and who is not. Boundaries can in general be defined by network members themselves (the realist approach) or by a researcher (the nominalist approach). The main distinction is the conscious experience of network membership that is required by the realist approach. In general, the realist approach works better for small groups and for very obvious activities and events (e.g., school class, participants in an accident), whereas the nominalist approach is better for larger, more formal groups (e.g., formal organizations, elites). When membership of the whole network is established, measurement of social networks is similar as in ego-centered networks (e.g., by name generators).

Free Recall and Recognition

The second is the method of eliciting the names of network members. The respondents can be either asked to freely recall them from memory (recall method) or they can chose them from a list provided in advance (recognition method or roster). With ego-centered networks free recall is used almost exclusively, whereas with whole networks both methods are often used.

Name Interpreters

As already mentioned, after collecting the names of network members, usually some questions about these persons and relationships with them are asked. These questions are called name interpreters. These are ordinary survey questions, such as "What is the gender of person X?" or "How long (in years) have you known person X?." These questions can be asked in two ways. The first possibility (by alters or alter-wise) is to take each network member separately and ask all name interpreters for this person before moving on to next network members. The second possibility is to take each name interpreter separately (by question or question-wise) and ask the question for all alters before moving to the next name interpreter. The results regarding which of the two methods is better are inconclusive, but it seems that by alters method works better in telephone and face-to-face surveys and by questions method in web surveys (e.g., Kogovšek et al. 2002; Vehovar et al. 2008).

Future Directions

The expansion of computer-assisted questionnaires for social network data in general and web questionnaires in particular is a trend which is expected to continue in the future. In the context of the latter developing questionnaires for online communities where part of the user's network would be provided automatically (his/her online contacts, e.g., Facebook friends) and part would be asked could be a new challenge, as well as developing questionnaires for new technologies (e.g., smartphones), one trend may also be towards reducing respondent burden and data collection costs (e.g., balancing limitation of the number of elicited network members on the one hand and data quality on the other).

Cross-References

- ▶ Collection and Analysis of Relational Data in Organizational and Market Settings
- ▶ Network Data Collected via the Web
- ▶ Quality of Social Network Data

References

Bailey S, Marsden PV (1999) Interpretation and interview context: examining the general social survey name generator using cognitive methods. Soc Netw 21: 287–309

Burt RS (1984) Network items in the general social survey. Soc Netw 6:293–339

Coenders G, Doreian P (eds) (2011) Social networks and performance of PhD students. Spec Issue Metodološki Zvezki 8(2):115–207

Granovetter M (1973) The strength of weak ties. Am J Sociol 78:1360–1380

Hlebec V, Kogovšek T (2013) Different approaches to measure ego-centered social support networks: a meta-analysis. Qual Quant 47:3435–3455

Hogan B, Carrasco JA, Wellman B (2007) Visualizing personal networks: working with participant-aided sociograms. Field Methods 19:116–144

Holland PW, Leinhardt S (1973) The structural implications of measurement error in sociometry. J Math Sociol 3:85–111

Kahn RL, Antonucci TC (1980) Convoys over life course: attachment, roles and social support. In: Baltes PB, Brim OG (eds) Life-span development and behavior. Academic, New York, pp 253–286

Kogovšek T, Ferligoj A (2004) The quality of measurement of personal support subnetworks. Qual Quant 38:517–532

Kogovšek T, Ferligoj A (2005) Effects on reliability and validity of egocentered network measurements. Soc Netw 27:205–229

Kogovšek T, Hlebec V (2005) Effects of limitation of number of alters and time frame in the Burt name generator. Metodološki zvezki 2:59–71

Kogovšek T, Ferligoj A, Saris WS, Coenders G (2002) Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. Soc Netw 24: 1–20

Laumann EO, Marsden PV, Prensky D (1983) The boundary specification problem in network analysis. In: Burt R, Minor M (eds) Applied network analysis: a methodological introduction. Sage, Beverly Hills, pp 18–34

Lin N, Fu Y, Hsung R (2001) The position generator: measurement techniques for social capital. In: Lin N, Cook K, Burt RS (eds) Social capital: theory and research. Aldine, New York, pp 57–84

McCallister JL, Fischer CS (1978) A procedure for surveying personal networks. Sociol Methods Res 7: 131–148

McCarty C, Molina JL, Aguilar C, Rota L (2007) A comparison of social network mapping and personal network visualization. Field Methods 19: 145–162

Milardo RM (1989) Theoretical and methodological issues in the identification of the social networks of spouses. J Marriage Fam 51:165–174

Moreno JL (1934) Who shall survive? Nervous and Mental Disease Publishing Company, Washington, DC

Padgett JF, Ansell CK (1993) Robust action and the rise of the medici, 1400–1434. Am J Sociol 98: 1259–1319

Sampson SF (1968) A novitiate in a period of change: an experimental and case study of social relationships. Ph.D. thesis, Cornell University, Ithaca, NY

Van der Gaag M, Snijders TAB (2005) The resource generator: social capital quantification with concrete items. Soc Netw 27:1–29

Van der Poel MGM (1993) Delineating personal support networks. Soc Netw 15:49–70

Van Sonderen E, Ormel J, Brilman E, Van Linden van der Heuvell C (1990) Personal network delineation: a comparison of the exchange, affective and role relation approach. In: Knipscheer CPM, Antonucci TC (eds) Social network research: substantive issues and methodological questions. Swets and Zeitlinger, Amsterdam, pp 101–120

Vaux A (1988) Social support: theory, research, and intervention. Praeger, New York

Vehovar V, LozarManfreda K, Koren G, Hlebec V (2008) measuring ego-centered networks on the web: questionnaire design issues. Soc Netw 30: 213–222

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

---

Queueing Theory

Myron Hlynka
Department of Mathematics & Statistics,
University of Windsor, Windsor, ON, Canada

Synonyms

Service systems; Waiting time models

Glossary

Bottleneck A location in a queueing network where the delay is especially long, perhaps due to a reduction in service at that location

Busy Period For a single server system, the busy period is the time between the arrival of a customer to an empty system until the server has no customers remaining to be served

Call Center A telephone service center manned by numerous servers in which "customers" can phone for information or service

Customer A person or object that must receive service. Examples could be people, cars, computer commands

Double Ended Queue A system with a line of customers or a line of servers and the interaction between them (such as taxis and passengers)

First In First Out (FIFO)/First Come First Serve The procedure of customers joining the queue and being served in order corresponding to the arrival time

Interarrival Time The time between two consecutive arrivals to the system

Jockeying The procedure of a customer switching between multiple queues, (usually) attempting to shorten the system

M/G/1 Queue A queueing system with exponentially distributed interarrival times and arbitrarily distributed service times

M/M/1 Queue A queueing system with exponentially distributed interarrival times and exponentially distributed service times, and a single server

Memoryless Property A statistical distribution has the memoryless property if conditional probabilities about the residual life (time since last arrival or remaining service time) at any selected time point is distributionally the same as the original lifetime. For continuous distributions, the exponential distribution uniquely has the memoryless property

Priority Queueing System A system in which customers are categorized into groups and where members of certain groups receive service before members of other groups. An example would be the triage system of a hospital emergency room

Queue A line of waiting "customers"

Queue Jumping The procedure of customers joining the queue somewhere other than at the end (sometimes considered cheating)

Queueing Network A sequence of queues in series or in parallel through which a customer may pass

Reneging The procedure of dropping out of a queue (usually) because of excessive waiting

Server A person or object that provides some kind of service. Examples could be dentists, car repair robots, computer programs

Service Time The time that a customer spends in service

Split and Match Queue A queueing system in which a customer is split into two parts which receive service separately and then must be matched back together in order to complete the service

System Time The sum of the waiting time and the service time. The system time is the total time that a customer spends in the system

Traffic Intensity This is a unitless measure of the how busy the server is in a queueing system

Waiting Time The time that a customer spends waiting before entering service

Definition

Queueing theory is the mathematical study of waiting lines. Those needing service are called customers and those giving service are called servers. If the service times were all known exactly ahead of time, and if the customers were punctual, it would be possible to have a system with no waiting time and no queueing. However, random times in either the interarrival times or the service times or poor scheduling will result in some kind of queueing. The word "queueing" is also spelled "queuing"; the former version is preferred by most researchers in the field.

Introduction

Queues (lineups) exist in many aspects of society. We wait in a queue when we go to a bank, or to a grocery store, or to a fast-food restaurant. We wait in a queue to board an airplane. We wait in a queue when we visit a medical clinic. We wait in a queue for rides at an amusement park. We wait in a queue to buy tickets for a concert. We wait in our vehicle for a green traffic signal. Queues do not apply just to people. We may have a queue of exams that must be graded, a queue of inventory that must be delivered, a queue of tasks that must be completed, a queue of e-mails that must be organized, or a list of text messages that must be read.

Queues can be controlled (provide more servers, change traffic light timing, lower the arrival rate by charging fees). Queueing times can be estimated ("Your estimated wait is five minutes.") and these estimates are valuable. Queues can be modified by adding priority service. The change in queueing times can be predicted when conditions change. Thus, queues are worthy of serious study.

Key Points

A simple classification method for queues has the form A/B/C where A represents the interarrival time distribution, B represents the service time distribution, and C represents the number of servers. The most common interarrival or service distribution notations are M (memoryless, exponential), D (deterministic), G (general), or GI (general input). Thus, we find thousands of papers in the research literature on M/M/1 or M/D/1 or M/G/1 or G/M/1 queues. This notation has been extended. An M/M/2/5/5 system, for example, refers to a queueing system with exponential interarrival times, exponential service times, 2 servers, a capacity for 5 customers, and a population from which the customers are selected of size 5. This could be used as a model for machine breakdown with a population of five machines and two servers.

Two commonly observed types of queueing systems are:

(a) Single-line, multiple-server systems
(b) Multiple-line, multiple-server systems

Most banks use single-line multiple-server systems. Most grocery stores use multiple-line systems, each line with its own server. If no jockeying is allowed, then the expected system time per customer is lower for the single-line multiple-server system. Also the variance of the system time is lower for the single-line multiple-server system. The single-line system is also regarded as being fairer to customers. Thus, it is the preferred system in many settings. A better way to lower the expected system time is to serve the customers who require shorter service first. This is the reason for express checkout systems in some locations.

Historical Background

The beginning of modern queueing theory research is generally attributed to a Danish telephone engineer, Agner Krarup Erlang, who in 1909 published his first paper (Erlang 1909).

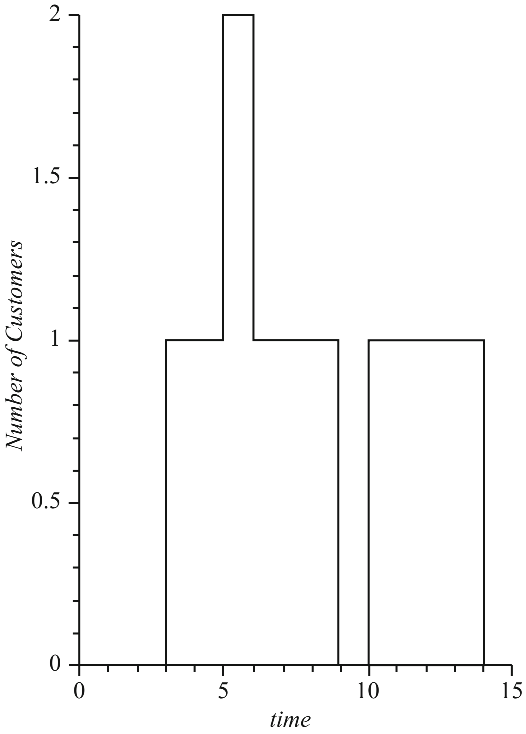
Tore Olaus Engset published his first queueing paper in 1918 (Engset 1918). Some of the other pioneers in queueing theory and dates of their first major work in queueing are O'Dell (1920), Molina (1922), Fry (1928), Pollaczek (1930), Kolmogorov (1931), Khinchin (1932), Crommelin (1932), and Palm (1936). A good summary of the history of queueing theory up to 1961 can be found in Saaty (1961). Classic texts prior to 1980 include work by Morse (1958); Cox and Smith (1961); Takacs (1962); Riordan (1962); Prabhu (1965); Lee (1966), Gross and Harris (1974) and (Kleinrock 1975, 1976).

Lindley (1952) presented a standard way of setting up integral equations to describe certain queueing systems. Kendall (1953) introduced the A/B/C type queueing notation. The concept of reneging (or queueing abandonment) was introduced by Palm (1937) and was rediscovered and named by Haight (1959). Haight (1957) also introduced the concepts of balking and parallel queues. Cox (1995) introduced the "supplementary variables technique" to analyze queues. Cohen (1957) was the first to consider retrial queues. White and Christie (1958) were the first to consider server breakdown (or vacation) queueing models. There are two individuals named Jackson with major contributions to queueing networks – R.R.P. (Ray) Jackson and James P. Jackson. R.R.P. Jackson (1954, 1956) and Jackson (1957, 1963) worked on queueing networks and product form solutions. Little (1961) gave a proof of Little's Theorem. Gaver (1968) and Newell (1968) introduced the idea of diffusion approximations for queues. Mandelbaum and Avi-Itzhak (1968) introduced the concept of the split-and-match (a.k.a. fork-join) queue. The use of queueing for computer performance evaluation began around 1970. Brill (1975) developed the level crossing method. Grassmann et al. (1985) developed the numerical GTH method. Cooper and Murray (1969) did some of the earliest work on polling systems. Neuts (1981) introduced the matrix analytic method. Wolff (1982) named, popularized, and gave a proof of the PASTA principle (Poisson Arrivals See Time Averages). Larson (1990) developed a queue inference engine. Gelembe (1991) introduced the concept of negative customers and G-networks.

Some major conferences in queueing include the matrix analytic conferences (began in 1995), the Canadian Queueing conferences (CanQueue, began in 1999), the Retrial Queue conferences (began in 1998), the QTNA (Queueing Theory and Network Applications) conferences, the MCQT (Madrid conference on Queueing Theory) series (began 2002), and BWWQT (Belarusian Winter Workshops in Queueing Theory) series (began 1985).

Foundations or Main Body

Queueing theory has an extensively developed mathematical framework. Let the states of the system represent the number of customers in the system. Then the number of customers in the system at time t depends on the number of



Queueing Theory, Fig. 1 Number of customers vs time

arrivals and the number of service completions up to time t.

The following diagram represents a typical graph of the number of customers versus time. There is an arrival at time 3, another arrival at time 5, a service completion at time 6, and another service completion at time 9. At time 10, there is an arrival which completes service at time 14 (Fig. 1).

Exponential Distribution-Based Models

The exponential distribution has the memoryless property, i.e., if X is exponentially distributed, then $P(X > a + b|X > a) = P(X > b)$.

Consider the M/M/1 queueing system, with exponentially distributed interarrival times and exponential service times, and a single server. This is the simplest system to analyze, because of the memoryless property. This means that to make probability statements about the future, we do not need to keep track of the time of the last service completion or the time since the current customer in service began its service. Changes in the number of customers occur if there are arrivals or service completions. Let $p_{i,j}(t)$ be the probability that the system changes from i customers in the system to j customers a time interval of length t. Let h be the length of a small time interval. Let λ be the arrival rate of customers and let μ be the service rate of customers. The number of arrivals in the interval $[0, t]$ has a Poisson distribution with mean λt. We will explain the following expression (for $i > 0$).

$$p_{ij}(t+h) = p_{i,j-1}(t)\lambda h + p_{i,j+1}(t)\mu h + p_{ij}(t)(1 - \lambda h - \mu h) + o(h). \tag{1}$$

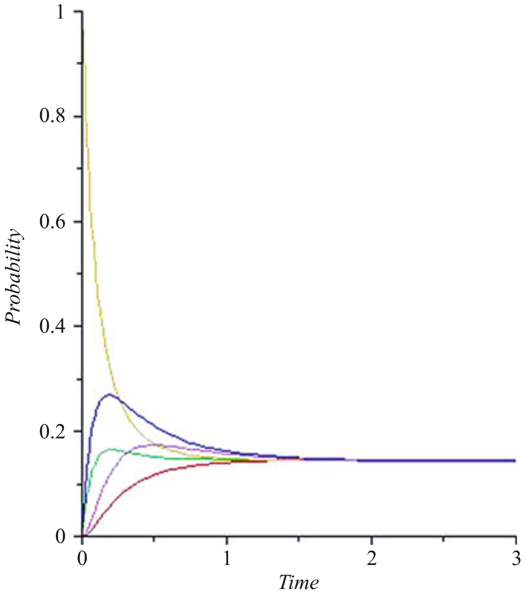
The left-hand side represents the probability of a change from state i to j in a time of length $t + h$. The right hand partitions this event into smaller pieces. There are three major ways to accomplish this change, which correspond to the first three summands on the right-hand side.

- The system can move from i to $j - 1$ in time t, with probability $p_{i,j-1}(t)$, and in the remaining small time h, exactly one customer may arrive with probability λh
- The system can move from i to $j + 1$ in time t, with probability $p_{i,j+1}(t)$, and in the remaining small time h, exactly one customer may complete service with probability μh.
- The system can move from i to j in time t, with probability $p_{i,j}(t)$, and in the remaining small time h, no customers arrive or complete service, with probability $1 - \lambda h - \mu h$.

The final summand on the right-hand side $o(h)$ represents any small probabilities that were not included in the first three summands. Here $o(h)$ is so small that

$$\lim_{h\to 0} \frac{o(h)}{h} = 0.$$

In (1), we subtract $p_{ij}(t)$ from both sides, divide by h, and let $h \to 0$. Then we recognize the new left-hand side as the definition of a derivative. The resulting system of differential equation is, for $i > 0$,



Queueing Theory, Fig. 2 Probability $p_{ij}(t)$ vs t *for* $i = 0, 1, 2, 3, 4$ and $j = 2$

$$\frac{d}{dt}p_{ij}(t) = p_{i,j-1}(t)\lambda + p_{i,j+1}(t)\mu + p_{ij}(t)(-\lambda - \mu). \tag{2}$$

It is remarkable that a relatively simple queueing model gives rise to a differential equation to describe the model. Using appropriate initial conditions, it is possible to solve for $p_{i,j}(t)$, although the solution is very complicated. The following graph shows $p_{i,j}(t)$ for j $= 2$, i $= 0, 1, 2, 3, 4$, when $\lambda = 3$ and $\mu = 5$.

In Fig. 2, note that all four of the curves tend to the same level as $t \to \infty$.

This indicates that $\lim_{t\to\infty} p_{ij}(t)$ does not depend on i, so we can define the limiting probabilities as $\pi_j = \lim_{t\to\infty} p_{ij}(t)$, for $j = 0, 1, 2, \ldots$. These limiting probabilities, also called steady-state probabilities, are very important, and π_j represents the long run proportion of time that there are j customers in the system. Thus, π_0 represents the long run proportion of time that the system is empty and if we let L represent the number of customers in the system, then $E(L) = \sum_{j=0}^{\infty} j\pi_j$ is the expected number of customers in the system. Not only does Fig. 2 allow us to define π_j but it also shows that $\lim_{t\to\infty} \frac{d}{dt}p_{ij}(t) = 0$.

Computing the Limiting Probabilities for an M/M/1 System

There can be $0, 1, 2, \ldots, j-1, j, j+1, \ldots$ customers in the queueing system. We can draw the transition diagram as follows (Fig. 3).

Since $\lim_{t\to\infty} \frac{d}{dt}p_{ij}(t) = 0$, (2) becomes, for $j > 0$

$$0 = \pi_{j-1}\lambda + \pi_{j+1}\mu + \pi_j(-\lambda - \mu). \tag{3}$$

This can be rewritten as

$$\pi_{j-1}\lambda + \pi_{j+1}\mu = \pi_j(\lambda + \mu). \tag{4}$$

If we focus on states $j-1, j, j+1$, we have the following partial transition diagram.

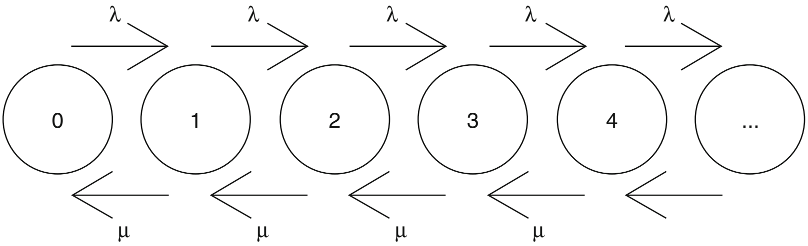
Looking at expression (4), we see that the left-hand side represents the rate of transitions into state j and the right-hand side represents the rate of transitions out of state j, for $j > 1$. With this interpretation, expression (4) becomes a *balance equation*. A related balance equation results by looking at two consecutive states $j-1$ and j as in Fig. 4. If we imagine a vertical line between the two states, then flow across the line in a right-hand direction should balance with the flow across the line in a left-hand direction. This gives us the equation

$$\pi_{j-1}\lambda = \pi_j\mu$$

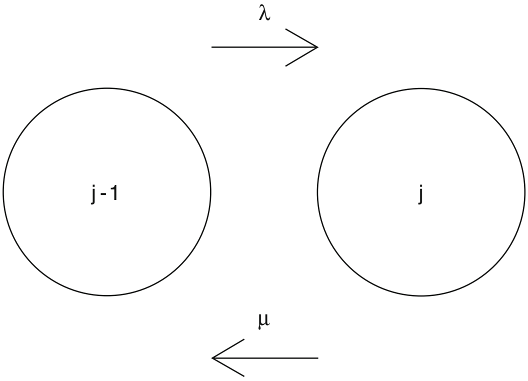
for $j = 0, 1, 2, \ldots$ Thus, we find

$$\pi_j = \frac{\lambda}{\mu}\pi_{j-1} = \left(\frac{\lambda}{\mu}\right)\frac{\lambda}{\mu}\pi_{j-2} = \cdots = \left(\frac{\lambda}{\mu}\right)^j \pi_0$$

Since the limiting probabilities should sum to 1, we have



Queueing Theory, Fig. 3 Transition rates between adjacent states



Queueing Theory, Fig. 4 Transitions between states $j-1$ and j

$$1=\sum\nolimits_{j=0}^{\infty}\pi_j=\sum\nolimits_{j=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^j\pi_0$$
$$=\frac{1}{1-\left(\frac{\lambda}{\mu}\right)}\pi_0$$

So $\pi_0=1-\frac{\lambda}{\mu}$, if $\frac{\lambda}{\mu}<1$.

We can summarize the above discussion by saying that if $\lambda<\mu$, for an M/M/1 queue, the steady-state probabilities that the system has j customers (for $j=0,1,\ldots$) are

$$\pi_j=\left(\frac{\lambda}{\mu}\right)^j(1-\frac{\lambda}{\mu}).$$

For an M/M/c system, define the traffic intensity as $\rho=\frac{\lambda}{c\mu}$, which is a measure of how busy the system is. In particular, if $c=1$, we have an M/M/1 system, and $\rho=\frac{\lambda}{\mu}$. Hence, for the M/M/1 system, $\pi_j=\rho^j(1-\rho)$ for $j=0,1,\ldots$.

We recognize this as the probability mass function of a geometric random variable. It is somewhat surprising to see the geometric distribution make its appearance here. The traffic intensity ρ is a unitless quantity, since both λ and μ have the same units and ρ is their ratio. In spite of this, ρ is sometimes stated in terms of unitless "erlangs."

Queueing systems of the M/M/* type have associated rate matrices (or infinitesimal generators). The rows all sum to zero so zero is an eigenvalue of every rate matrix. The states of the system represent the number of customers in the system. If we label the rows as 0,1,2, ... and the columns as 0,1,2,..., then the entry in the (i, j) position (row i and column j) then nondiagonal entries represents the rate at which the process is changing from state i to state j.

For example, an M/M/1 queueing system has rate matrix

$$Q=\begin{bmatrix}-\lambda & \lambda & 0 & 0\ \ldots\\ \mu & -(\lambda+\mu) & \lambda & 0\ \ldots\\ 0 & \mu & -(\lambda+\mu) & \lambda\ \ldots\\ \vdots & \vdots & \ddots & \ddots\ \ddots\end{bmatrix}.$$

Here the arrival rate is λ customers per unit time and the service rate is μ customers per unit time.

A social network group with five members and joining rate λ per unit time per member and exiting rate μ per unit time per member would have states 0, 1, 2, 3, 4, and 5 representing the number of people active in the system. A plausible rate matrix is

$$Q = \begin{bmatrix} -5\lambda & 5\lambda & 0 & 0 & 0 & 0 \\ \mu & -(4\lambda+\mu) & 4\lambda & 0 & 0 & 0 \\ 0 & 2\mu & -(3\lambda+2\mu) & 3\lambda & 0 & 0 \\ 0 & 0 & 3\mu & -(2\lambda+3\mu) & 2\lambda & 0 \\ 0 & 0 & 0 & 4\mu & -(\lambda+4\mu) & \lambda \\ 0 & 0 & 0 & 0 & 5\mu & -5\mu \end{bmatrix}.$$

Let $\vec{\pi} = (\pi_0, \pi_1, \pi_2, \cdots)$ be a row vector containing the limiting probabilities. Let $\vec{0} = (0, 0, \cdots)$ be the zero tow vector. Solving for $\vec{\pi}$ in the matrix equation $\vec{0} = \vec{\pi} Q$ with the condition $\sum_{j=0}^{\infty} \pi_j = 1$ yields the desired limiting probabilities. Here the choice of five members is arbitrary and can be replaced by an arbitrary value N.

In the above two examples, the rate matrices are tridiagonal (or both upper Hessenberg and lower Hessenberg). This is the case for many (not all) queueing models.

Transition Matrices

Although rate matrices can be used to describe some models, they are not always appropriate. In the M/G/1 system, for example, the standard method is to look at the system immediately after a service completion. The states of the system represent the number of customers present, namely, 0, 1, 2, … . Corresponding to these states, we create a transition matrix P with rows numbered 0, 1, 2, … and columns number 0, 1, 2, …. Entry p_{ij} represents the probability of a change from i customers just after the most recent service completion to j customers just after the end of the next service completion. Thus, a drop in the number of customers can be of size at most 1, whereas the increase can be arbitrarily large. For example, if there are three customers in the system and one is being served, and if there are four arrivals before the service is completed, then the transition is from 3 to 6 since just after the completion there will remain $3 - 1 + 4 = 6$ customers.

As a result, the transition matrix has the appearance

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

A transition from 0 to 2, for example, means that an arrival must occur and that during the service of that first arrival, there must be two other arrivals with probability a_2. A transition from 1 to 2 means that the last service completion left 1 customer in the system and during the service of the one customer, there are two arrivals with probability a_2. Thus, the first two rows are identical. It turns out that the transition matrix can be used to obtain the limiting probabilities by solving $\vec{\pi} = \vec{\pi} P$, subject to $\sum_{j=1}^{\infty} \pi_j = 1$.

Simulation Methods

Many queueing applications experts will use simulation methods to analyze queueing systems. They argue that using exponential based models is not realistic since most systems have arrival or service distributions that are not exponential and the quantities that need to be measured are strongly affected by the underlying model. The simulation experts may be able to observe the system and find an accurate model which can be simulated. Then they can change the structure or parameters of the system to determine the effect of these changes in an accurate and useful way. Simulation models can be developed much faster than analytic models and lend themselves to visual presentation.

Other queueing experts like analytic mathematical models. They point out that the exponential interarrival model is actually quite common in practice since it results when customers behave independently of each other, which is often the case. Further, they can build a network of exponential type models which can approximate any system, by using phase type distributions and matrix analytic methods. They argue that the generality of their analytic models allows them to see insights into the operation of the queueing system and to see its limitations more clearly.

Most queueing theorists are open to any method, analytic or simulation, that will allow them to get valuable results to describe, understand, predict, and control queueing systems.

More General Models

A widely used result in queueing theory is Little's Theorem. This result is easy to state and is true for surprisingly general conditions.

Let L represent the length (or the number of customers) in the queueing system. Let E(L) represent the expected number of customers in the system (average over time).

Let W represent the wait (actually the system time) of a randomly arriving customer to the system, and let E(W) be the expected system.

Let λ and μ represent the arrival and service rates, respectively.

Little's Theorem

For a GI/G/c queueing system,

$$E(L) = \lambda E(W).$$

Another popular result is the PASTA principle (Poisson Arrivals See Time Averages).

If customers arrive according to a Poisson process, the expected value of system characteristics viewed by these customers is the same as the time average of the system.

This result is valuable in studying M/G/c systems.

We can also use the PASTA principle together with Little's Theorem to get valuable information. For example, suppose we have an M/M/1 queueing system, where L represents the number of customers in the system.

By Little's Theorem, $E(L) = \lambda E(W)$.

By the PASTA principle, a randomly arriving customer will see j customers with probability π_j. By the memoryless property of the exponential distribution, each of the j customers (plus the randomly arriving customer) has a full service time left. Let $E(W|j)$ be the time to finish j customers. The expected service time for a single customer is the reciprocal of the service rate, namely, $\frac{1}{\mu}$. Thus, using the fact that $\sum_{j=0}^{\infty} \pi_j = 1$, we have

$$\begin{aligned} E(W) &= \sum\nolimits_{j=0}^{\infty} \pi_j E(W|1+j) \\ &= \sum\nolimits_{j=0}^{\infty} \pi_j \frac{1+j}{\mu} \\ &= \frac{1}{\mu}\left(1 + \sum\nolimits_{j=0}^{\infty} \pi_j j\right) \\ &= \frac{1}{\mu}(1 + E(L)). \end{aligned}$$

Combining information gives $E(L) = \lambda E(W) = \lambda \frac{1}{\mu}(1 + E(L)) = \rho(1 + E(L))$.

Solving gives

$$E(L) = \frac{\rho}{1-\rho}.$$

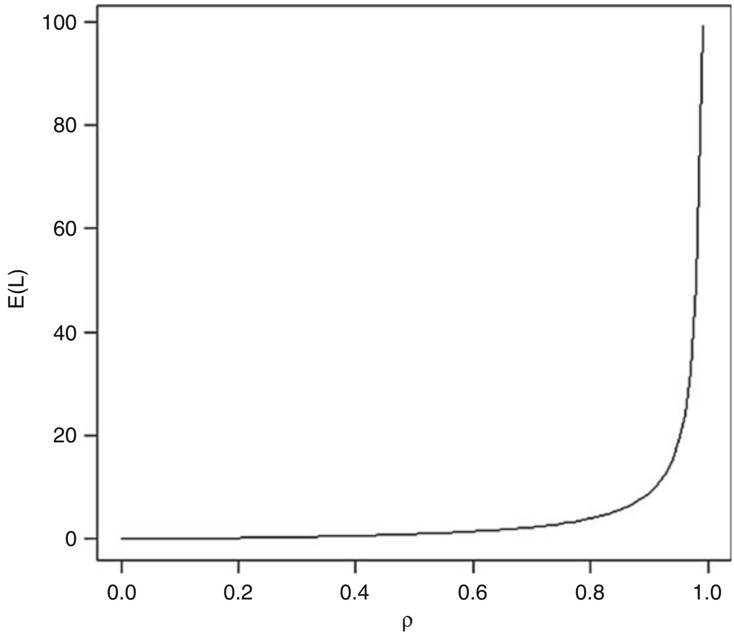
Stability Condition

It is easy to understand that if the queueing system is D/D/1, and if the arrival rate λ exceeds the service rate μ, then the queueing system will continue to grow and system is stable if and only if $\lambda \leq \mu$ or $\rho \leq 1$.

However, for an M/G/1 queueing system, the stability condition is $\lambda < \mu$ or $\rho < 1$, which is slightly different. As ρ approaches 1, E(L) approaches ∞. This can be illustrated in the following diagram, which was based on an M/M/1 system (Fig. 5).

The key lesson to be learned from this is that one should never assume that as long as the average arrival rate is less than the average service rate, the queue length (and the wait times)

Queueing Theory, Fig. 5 Expected length (number of customers in system) vs ρ



will be reasonable. Under most circumstances, the arrival rate/service rate ratio should never be allowed to approach too close to 1. The stability condition for an M/M/s system is still $\rho < 1$ where $\rho = \lambda / s\mu$.

The fact that the distribution of the number of the customers in the system (and the expected number of customers in the system) only need the knowledge of ρ and do not require the knowledge of both λ and μ leads to some interesting observations.

The Camcorder Principle

(a) Certain measures of queueing systems (such as expected system length or proportion of time that the system has i customers) will not change if the units of time change. The camcorder principle says that such measures will be functions of ρ and do not require the knowledge of both λ and μ.

(b) Other measures of queueing systems (such as the time in the system or the time length of a busy period) decrease as the time units get larger.

Intuitively, if we made a video of a queueing system, the proportion of time the system has i customers would be the same whether we played the video at regular of high or low speed. But the arrival rate and the service rate would change proportionately so the traffic intensity ρ would not change. So we would expect our formula for E(L) to involve only ρ, whereas E(W) would be expected to involve both λ and μ (or both λ and ρ or both μ and ρ).

Another classic result is stated below.

Pollaczek-Khinchin Formula

For an M/G/1 queue, let E(L) be the expected number of customers in the system. Let S be the random variable for the service times. Then

$$E(L) = \rho + \frac{\rho^2 + E(S)}{2(1 - \rho)}.$$

Erlang Formulas

There are two important results in queueing which are referred to as Erlang B and Erlang C. Erlang B gives the probability that a customer arriving to an M/G/s/s queueing system (with s servers and capacity s) will be lost because the system is full. This result, also called the Erlang loss formula, is important in telecommunication systems. The Erlang C formula finds the probability that an arriving customer will have to wait before entering service in an M/M/c model. There are an enormous number of online Erlang calculators which are used to determine the number of servers (staffing levels) needed to attain a high probability of a short specified wait for customers in call center settings.

Illustrative Example(s)

We wait in a queue when go to an emergency waiting room at a hospital. In this case, we patients may receive a priority status assigned by a triage nurse. Other priority queues occur in the post office, where we can pay extra for faster service for our packages.

Some queues are not so visible as we may not be able to see other customers. We are assigned to a waiting list for a medical operation, which may be months away. We wait for service for home repair, which may be far in the future. We wait for a date for a court trial. Students may be on a waiting list to get into a section of a course which is currently filled. In such cases, "customers" can continue other activities while they are waiting. In a call center, there may be many people waiting for service.

Most queues do not involve people. We may have an assembly line of items that need to be processed. We may have a set of computer commands that need to be executed, perhaps in some particular order. In telecommunications, there may be information packets, which must be queued and processed.

There are double-ended queues. The canonical example consists of taxis and passengers. We may have a line of taxis waiting for passengers or we may have a line of passengers waiting for taxis. We can view this as the excess of taxis over passengers where the "excess" may be positive of negative. At blood banks, we may have a supply of blood waiting for patients or a group of patients waiting for blood. The same is true for organ transplants (liver, heart, eyes, kidneys). We can view sporting competitions as double-ended queues. The excess of points of team A over two B may be positive or negative. Similarly, a military conflict can be viewed as a double-ended queue. In trading stocks on the financial market, there is a line of buyers and sellers who are trying to agree on acceptable prices.

Sometimes a system may not be recognizable as a queue. Cars arrive at a parking lot. Each car can be considered to be a customer and each parking spot may be considered to be a server so there is always a server for every accepted customer, although there is a limit on the number of customers who can enter the system. In the parking lot example, we may be interested in where the cars park or we may only be interested in the total number of cars which are in the parking lot.

Consider a social network of students in a university course who can enter a site in which they can discuss the course, homework problems, and related issues. This is very much like the parking lot. Students enter the system, stay for a while, and then choose to leave, perhaps to return later. We might be interested in how many students are in the system at various time points. A set of food items in a grocery might be considered to be customers waiting for servers (purchasers) to serve (purchase) them.

Queueing networks are formed from a chain of queues. Queues may be connected in series or in parallel. It is possible to have customers in the network at various stages. An assembly line may be viewed as a manufacturing queueing system.

Key Applications

Queueing theory initially was used to determine the number of trunk lines for telephone systems. Later applications involved communication systems and digital routers. Applications have also looked at manufacturing systems and assembly lines and scheduling. Queueing has been used to study performance analysis in computers and in computing networks. Users are often surprised that a major increase in the speed of a computer processor in one position may result in no significant performance improvement, due to the fact that the system time is mainly dependent on a bottleneck at some other location in the system. Queueing modes have been used to improve health-care service including emergency services. They have been used for entertainment parks and retail sales (How many cash registers are needed? How many dressing rooms are needed?). Over the years, banks, fast-food restaurants, and grocery stores have modified and improved their queueing systems. Queueing models are being used to analyze and describe call centers.

Queueing models are important in vehicle traffic flow. The position and duration of traffic lights, the staggered work day of employees, the speed limit on highways, all affect the amount of traffic that can travel through the system and the time needed to traverse the network.

Future Directions

Current popular techniques include fluid models for queues (Gautam 2012), matrix analytic techniques (Latouche and Ramaswami 1999), simulation methods (Stewart 2009), level crossing methods (Brill 2008), and discrete queueing methods (Alfa 2010). Popular topics include applications in health care, priority queueing systems, communications systems, computer networks, call centers, cell phone operations, and transient queues (measured at a particular finite time point from the initial start).

Queueing models are important in traffic of any kind. Internet traffic and speed and capacity are improving rapidly with the help of technology in response to rapidly changing demands. Cloud computing is increasing in popularity. Networks such as FACEBOOK or YouTube or GOOGLE require special attention. Each change requires new or modified queueing analysis.

Cross-References

- ▶ Distributed Processing of Networked Data
- ▶ Exchange Networks
- ▶ Matrix Algebra, Basics of
- ▶ Network Models
- ▶ Probabilistic Analysis
- ▶ Probability Matrices
- ▶ Simulation
- ▶ Spectral Analysis
- ▶ Theory of Probability, Basics and Fundamentals

References

Alfa A (2010) Queueing theory for telecommunications. Springer, New York

Bolch G, Greiner S, de Meer H, Trivedi KS (2006) Queueing networks and markov chains, 2nd edn. Wiley-Interscience, Hoboken

Brill PH (1975) System point theory in exponential queues. PhD thesis, Industrial Engineering, University of Toronto

Brill P (2008) Level crossing methods in stochastic processes. Springer, New York

Cohen JW (1957) Basic problems of telephone traffic theory and the influence of repeated calls. Philips Telecommun Rev 18:49–100

Cooper R, Murray G (1969) Queues served in cyclic order. Bell Syst Tech J 48:675–689

Cox DR (1955) The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. Proc Camb Philos Soc 51:433–441

Cox DR, Smith WL (1961) Queues. Methuen, London

Crommelin CD (1932) Delay probability formulae when the holding times are constant. Post Off Electr Eng J 25:41–50

Engset T (1918) Die Wahrscheinlichkeitsrechnungzur Bestimmung der Wahlerzahl in automatischen Fernsprechamtern. ElektrotechnischeZeitschrift 31:304–306

Erlang AK (1909) The theory of probability and telephone conversations. (in Danish). NytTidsskrift for Matematik B 20:33–39

Fry T (1928) Probability and its engineering uses. Van Nostrand, New York

Gautam N (2012) Analysis of queues: methods and applications. CRC, Boca Raton

Gaver DP (1968) Diffusion approximations and models for certain congestion problems. J Appl Probab 5: 607–623

Gelembe E (1991) Product form queueing networks with negative and positive customers. J Appl Probab 28:656–663

Grassmann WK, Taksar M, Heyman DP (1985) Regenerative analysis and steady state distributions for Markov chains. Oper Res 33:1107–1117

Gross D, Harris CM (1974) Fundamentals of queueing theory. Wiley, New York

Gross D, Shortle JF, Thompson JM, Harris CM (2008) Fundamentals of queueing theory, 4th edn. Wiley, Hoboken

Haight FA (1957) Queuing with balking. Biometrika 44: 360–369

Haight FA (1959) Queueing with reneging, Metríka 2: 186–197

Hlynka M (2013) Queueing theory web site. http://web2.uwindsor.ca/math/hlynka/queue.html. Accessed 15 May 2013

Jackson RRP (1954) Queuing systems with phase type service. Oper Res Quart 5:109–120

Jackson RRP (1956) Random queuing processes with phase type service. J R Stat Soc B 18:129–132

Jackson JR (1957) Networks of waiting lines. Oper Res 5:518–521

Jackson JR (1963) Jobshop-like queuing systems. Manag Sci 10:131–142

Kendall DG (1953) Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chains. Ann Math Stat 24:338–354

Khinchin A (1932) Mathematisches uber die Erwartung vor einem offentlichen Schalter. Matematicheskii Sbornik 39:73–84

Kleinrock L (1975, 1976) Queueing systems. vols 1, 2. Wiley, New York

Kolmogorov AN (1931) Sur le probleme d'attente, Matematicheskii Sbornik 38:101–106

Larson RC (1990) The queue inference machine. Manag Sci 36:586–601

Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic modeling. SIAM, Philadelphia

Lee AM (1966) Applied queueing theory. MacMillan, London

Lindley DV (1952) The theory of queues with a single server. Math Proc Camb Philos Soc 48:277–289

Little JDC (1961) A proof of the queuing formula $L = \lambda W$. Oper Res 9:383–387

Mandelbaum M, Avi-Itzhak B (1968) Introduction to queueing with splitting and matching. Isr J Technol 6:288–298

Molina EC (1922) The theory of probabilities applied to telephone trunkingproblems. Bell Syst Tech J 1:69–81

Morse PM (1958) Queues, inventories, and maintenance. Wiley, New York

Neuts MF (1981) Matrix-geometric solutions in stochastic models. Johns Hopkins University Press, Baltimore

Newell GF (1968) Queues with time-dependent arrival rates. J Appl Probab 5:436–451

O'Dell GF (1920) Theoretical principles of the traffic capacity of automatic switches. Post Off Electical Eng J 13:209

Palm C (1936) Calcul exact de la perte dans les groupes de circuits echelonnes. Ericsson Tech 3: 41–71

Palm C (1937) Etudes des delais d'attente. Ericsson Tech 2:39–56

Pollaczek F (1930) Uber eine Aufgabe der Wahrscheinlichkeitstheorie. MathematischeZeitschrift 32: 64–100

Prabhu NU (1965) Queues and inventories. Wiley, New York

Riordan J (1962) Stochastic service systems. Wiley, New York

Saaty T (1961) Elements of queueing theory. McGraw-Hill, New York

Stewart WJ (2009) Probability, markov chains, queues, and simulation. Princeton University Press, Princeton

Takacs L (1962) Introduction to the theory of queues. Oxford University Press, Oxford

White H, Christie LS (1958) Queuing with preemptive priorities or with breakdown. Oper Res 6:79–95

Wolff RW (1982) Poisson arrivals see time averages. Oper Res 30:223–231