Using Imperfect Information in Online Social Systems: Applications and Measures of Betweenness

Miguel-Angel Sicilia
University of Alcalá
msicilia@uah.es

Elena García-Barriocanal
University of Alcalá
elena.garciab@uah.es

Nikolaos Korfiatis
Copenhagen Business School
nk.inf@cbs.dk

Abstract

Online social software systems enable individuals to build representations of their network of social ties, and such information can be used for the analysis of network structure and the position of individuals in the network. However, social ties representing relationships as "friendship" have intensity or strength as an intrinsic, essential feature, and this needs to be taken into account in network models. This paper reports on fuzzy techniques that deal with imprecise information on the strength of social ties, and provides an extended model of brokerage that accounts for such imprecise information.
Keywords: Social networks, social software, fuzzy numbers, betweeenness.

1 Introduction

Undoubtedly the use of the internet as a medium for social communication has paved new directions on the way people interact with each other and coordinate their social activities. Current social software systems on the Web (e.g. `LinkedIn`[1] or `Facebook`[2]) serve as a medium for people to register their social connections. This reflects a vast amount of dyadic relations that are subject to a wide range of common social contexts (e.g. a workplace, an organization or a team) as well as experiences (e.g. traveling together or met randomly in an event). Concretely, the registering of social links is done by request of the users, which count on profile search tools to find people they already know. In consequence, the formation of networks is driven by the interest of the users (usually in chained sequences of invitations), and by their subjective impression on who are his/her 'contacts' or 'friends'. However the intention of users of these platforms to become more visible on these social systems has led also to a *spam* phenomenon where users request to establish ties with users that are unknown to them only to increase their visibility in the online social system. However its up to the users to filter requests from strangers and there also several tools created to fight against this phenomenon [10].

This aspect of Social software is evidently useful for individuals as a way of recoding, maintaining and eventually strengthening their social ties. But it also provides a particular way to collect social network data that opens the opportunity to automatically compute diverse social network measures, which might be meaningful for analysis and decision making inside organizations [1], or for professionals and entrepreneurs using on-line tools. For example let us consider the case of a job market. With the difficulty of finding qualified professionals for specialized sectors such as IT, this wave of social systems is transforming the way the labor market works by adding new ways

[1]http://www.linkedin.com
[2]http://www.facebook.com/

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1105–1112
Torremolinos (Málaga), June 22–27, 2008

of assessing an individual's competence level e.g. by affiliation to trusted authorities.

The second wave in on-line social systems allows for creating *plugins* on top of established social platforms that are able to retrieve social information (restricted to the privacy policy of each individual). For example, in `Facebook`, `FQL` (*Facebook Query Language*) can be used to get the *userids* of the friends of a given individual, or the explicitly formed groups to which that individual belongs. This can be used in repeated queries to build a network representation in the form of a *undirected*, *nonvalued* graph that could be later used as input for social network software as `Pajek`[3]. However, this straightforward approach does not make use neither of subjective tie strength data nor of potential similarity sources (e.g. similarities on ratings for movies or books and the like). The strength of a social tie is an important element of the process of social network analysis due to the fact that interpretation of the results deriving from network models might be different due to factual strength of the tie that is taken into consideration. For example let us consider an analysis of network positions of individuals in a simple contact network. An analysis based on simple prestige measure might be misleading due to the fact that the strength of the tie is not considered [5]. The notion of the strength of a tie from a sociological perspective is a fundamental one [3]. In fact is an important element of social network analysis since it can affect the interpretation of fundamental concept of network analysis related e.g with prestige, centrality or network positions. For example in the simple case of prestige where the number of connections provides an indication of importance, the strength of adjacent network to the individual might affect the interpretation of the results.

The strength or the weakness of an individual tie however requires extensive knowledge of the individual's social environment which is sensitive under several different sources of social information. The use of these latter sources can lead to the creation of valued models, in which partial, subjective information could be combined to come up with better measures on the structural position of individuals. It should be noted the degrees of subjectively perceived strength are inherent to the dynamics of the relationships [16], so the use of representations for them come as a natural extension to enhance current social systems on the Web (and from a software development perspective, such extension poses no significant challenges).

[3]http://vlado.fmf.uni-lj.si/pub/networks/pajek/

This paper examines the current prospects for such kind of valued models based on imperfect information. Concretely, it focuses on a loose definition of "friendship" and sketches how social net measures could be derived from them. Since existing empirical evidence [9] point out that social systems on the Web contain strong local clusters, it seems reasonable to examine the individuals acting as brokers between the clusters as potential sources of communication and contact that add value [1] to the membership in on-line social systems.

Departing from this point the rest of this paper is structured as follows. Section 2 discusses how subjective assessments can be used and modeled in existing on-line social systems. Then, Section 3 deals with the use of that information for the analysis of structural holes. Finally, conclusions and outlook are provided in Section 4.

2 Modeling of Subjective Tie Strength

Since people in current social systems accept the registering of connections by themselves, it is reasonable to consider that all the ties have a positive valence. However, tie strength should be connected to a concrete interpretation of the meaning of the link. We focus here on "friendship", understood in a loose sense as a relationship that denotes co–operative and supportive behavior between two or more humans, be it actual or potential. Sites as `facebook` label as 'friends' to the people directly connected to one individual, however such notion is blurred with others as exchange relationships that are in use in the literature on the topic [13].

Obtaining objective measures or indicators of friendship is a challenge in itself, and it is not feasible with the information available in on-line social software systems as of today. Nonetheless, friendship relationships are potentially measurable since they represent time spent in socio–emotional interaction, but they are also subject to a continuous flux with dynamics that depend on several other different aspects [16]. This is offcourse due to the simplicity of the interfaces provided by the online social spaces which care mostly about user acquisition and retainment rather than providing the users tools for assessing their registered relations. Past or present co-working or co-participation in events then could be used also as an indicator, but social systems on the Web register only a small fraction of these relationships. As a consequence, the usefulness of general network models that use objective interaction as measure are limited in their reach of significant portions of the networks. A substitute for this kind of model–based measures is that of using straightforward subjective assessments of the strength of the tie.

Arguably the most common method of collecting data on personal and social networks is to ask people to recall network ties of one type or another [8]. The effort for an individual building his network in providing something as a subjective rating for the strength of his direct ties is not too onerous, especially if we consider that humans have a limited "friendship bandwidth" [15], which is consistent with the average figures of links found in empirical studies, e.g. in [9]. These figures range from less than five to slightly more than one hundred. Following that direction, this provides that an individual can provide a a subjective ordering of his/her important contacts which however would be subjective due to uncertainty derived from factors such as the psychological recall of the social *moments* experienced with these individuals. Even though strength assessment data is not being collected in some of the most popular online social systems today, it could be added to *plugin*-based sites by a simple extension. In the section that follows, we deal with the modeling of these assessments and their implication for network measures.

A second approach to build valued relationship networks is that of using some available indicator that is backed by some existing evidence, as preferences or tastes. For example, recent studies suggest that similarity in music preferences is related to friendship formation [12]. This results in different models, and we will not deal with them here.

2.1 Extending the Basic Social Network Model

A formal representation of the case discussed above requires us to the extend the standard social network model where people are represented as nodes in a graph and their social connections as edges. We will start from the common model of a (directed) graph $G = (P, A)$ where P is a nonempty finite set of n people (actors in general) and A defines a irreflexive relation that represent the *friendship assessments* of the people in P in terms of pairs of persons (arcs) $(x_i, x_j) \in A$. Then, values are introduced as a mapping S from the Cartesian product $P \times P$ to some *assessment domain* $\mathcal{A}$. So s_{ij} represents the closeness (friendship) assessment for the relation from actor x_i to actor x_j (in general, $s_{ij} \neq s_{ji}$), since the assessment may be asymmetrical.

An important data hiding principle is that the assessments of an individual cannot be disclosed to others, since the opposite would compromise the lack of bias of the subjective assessments. This affects applications but not the model in itself.

2.2 Direct Assessments of Friendship

The direct assessment of friendship can be done either by ordering friends or by stating a concrete linguistic label to describe the strength of the relationship. The latter approach can be mapped to fuzzy numbers, and linguistic label sets can be used for modeling scales resembling Likert ones. The model presented here is basic since it abstracts out factors as sex, age, economical status or race that are known to affect friendship, e.g. [14]. If information on those factors is available, arithmetics could be restricted to homogeneous groups.

Fuzzy numbers are fuzzy subsets of the real domain, and here we will consider only triangular fuzzy numbers, even though other models could be used instead. A triangular fuzzy number is represented by $\tilde{a} = (m, \alpha, \beta)$, with the membership function $\mu_{\tilde{a}}(x)$ is defined by the expression:

$$\mu_{\tilde{a}}(x) = \begin{cases} 0, & \text{if } x \le m - \alpha, \\ \frac{x-(m-\alpha)}{\alpha}, & \text{if } m - \alpha < x < m, \\ 1, & \text{if } x = m, \\ \frac{(m+\beta)-x}{\beta}, & \text{if } m < x < m + \beta, \\ 0, & \text{if } x \ge m + \beta, \end{cases} \tag{1}$$

where m is the center, α is the left spread and β is the right spread. Following the network model above, the assessment domain $\mathcal{A}$ will be that of these triangular fuzzy numbers $\mathcal{F}$.

Fuzzy numbers are able to capture uncertainty through the spread of the function, which is a useful tool for decision making in on-line Web interfaces. Also hedges and default models as *about* or *roughly*, or distance comparators as *much closer than* can be used to avoid the exposure of the modeling function to users, which would deal only with linguistic representations. These can be subject to empirical validation also.

From an egocentric perspective, this model allows for a number of applications, including the following:

i Controlling or sorting incoming information flows.

ii Regulating the automatic finding of potential friends by transitivity.

iii Having more realistic measures of central position.

Application [i] requires using one of the approaches for ordering fuzzy numbers. The following Table reports on a small exploratory study on using the approaches of Yager's F2 [17] and Liou and Wang [7]. Both numerical ranks were used on a data set of assessments for the egocentric networks of fifteen users, and the perception of appropriateness was elicited by showing metric distances on their contacts.

	Yager's F1	Liou and Wang
F(first option)	2	9
F(second option)	4	3
F(indiferent)	2	2

Since the index presented by Liou and Wang extends the one presented by Yager, the hypothesis was that it would be considered more appropriate for our approach. The estimation of triangular numbers was done by asking for ratings in a scale from zero to ten in the form (w, l, b), as the worse, more likely and better average for each contact. The findings confirm the hypothesis, so what is worth studying is the fuzzy number ranking methods that could better reflect the ranking behavior of individuals, or allowing on-line systems to dynamically change that behavior.

Application [ii] is usually done by checking friends-of-a-friend. Fuzzy arithmetics can be used to filter out some transitive possibilities in which the aggregated strength of the ties is slow or subject to a large degree of uncertainty. Other approaches might rank links in the network [6].

2.3 Fuzzy centralization measures

Application [iii] above entails extending measures of centrality. Here we will deal only with the extension of the in-degree centrality. As defined by Freeman [2], degree centrality is a count of the number of edges incident upon a given node (considering both directed and non-directed networks). Expression 2 is the usual matrix notation for degree centrality, and 3 is the standardized variant.

$$c_i^{DEG} = \sum_j a_{ji} \tag{2}$$

$$c_i^{SDEG} = \frac{\sum_j a_{ji}}{n-1} \quad (3)$$

The same degree centrality expression (2) applies in the case of a valued (weighted) social network with non negative numbers as tie values. However, for that case the standardization should be adjusted to the maximum possible aggregated value of the arcs adjacent to an actor, which entails changing the denominator in (3) to $(n-1) \cdot max$.

Expression 3 is equivalent to $c_i^{DEG} = \sum_j s_{ji}$ following our notation above, assuming that we count the incoming arcs to actor x_i. The extension of the model to the case of a fuzzy structural interpretation is straightforward, resulting in expression 4.

$$\widetilde{c_i}^{SDEG} = \frac{\sum_j \widetilde{s_{ji}}}{(n-1) \cdot \widetilde{max}} \quad (4)$$

Since each $\widetilde{s_{ji}}$ is a fuzzy number, the results of aggregating all the tie intensities will reflect the overall magnitude and the accumulated degree of uncertainty. Then, filters or rearrangement can be implemented related to these two aspects.

The reordering formula used as a case study was the index:

$$F_1(\mu(x)) = \frac{\int_{m-\alpha}^{\alpha} x \cdot \mu(x) dx}{\int_{m-\alpha}^{\alpha} \mu(x) dx} \quad (5)$$

For example, given $a1 = (3.8, 3, 1)$ and $a2 = (3, 0.1, 0.8)$, we have $F_1(a1) = 2.86$ and $F_1(a2) = 2.96$. The accumulated uncertainty for $a1$ results in a correction of the overall centrality score. In consequence, if we use $F1(\widetilde{c_i}^{SDEG})$ as the centrality measure, the results will be adjusted to the shape of the numbers used to assess the ties.

The rationale of the index is that of considering the area of the fuzzy number in the left slope, that is, the uncertainty related to lower centrality measures. For example, the number with $m = 3.8$ is considered to have an excessive degree of aggregated uncertainty for the left slope, so it is considered dubious as a high degree centrality.

3 Finding the actors that fill structural holes

Existing studies on social capital [1] discuss two main arguments on the network structures that create social capital. The *network closure argument* is that social capital is created by networks of strongly connected individuals. The *structural hole* argument is that social capital is created by a network in which some individuals can take the role of brokers between otherwise disconnected subnetworks. Recent studies [9] have reported that the graphs structure on common social software sites have a densely connected core comprising of between 1% and 10% of the highest degree nodes, such that removing this core completely disconnects the graph, but there is also evidence of the presence of strong local clustering. This suggests that the identification of the individuals acting as brokers for different segments are key in the creation of information flows that cross the boundaries of close friendship groups. These individuals are more likely to bring value to the contact system since they enable new relationships to be built and they can act as catalysts for information transfer across groups.

3.1 Extending Betweenness Measures

A well known measure for betweeenness centrality is Freeman's [2]. Betweenness centrality is defined as the share of times that an actor i needs another actor k (whose centrality is being measured) in order to reach a node j via the shortest path. A typical formulation of betweenness for actor k is as follows.

$$b_k = \sum_i \sum_j \frac{g_{ijk}}{g_{ij}} \; i \neq j \neq k \quad (6)$$

Where g_{ij} is the number of geodesic paths from i to j, and g_{ikj} is the number of these geodesics that pass through node k.

Fuzzy numbers can be used for modeling the problem, using for network measures one of the algorithms that solve the *fuzzy shortest path problem* (FSPP) [4].

A straightforward extension of the presented betweenness measures is that of considering the geodesics (shortest paths based on fuzzy numbers as values in the ties). Then, the above same measure can be used, and the difference lays in the use of fuzzy numbers to compute the geodesics. We will define $\widehat{g}_{ij}$ is the number of fuzzy geodesic paths from i to j, and $\widehat{g}_{ikj}$ as the number of these geodesics that pass through node k using fuzzy numbers for finding the shortest path.

However, this does not account for fuzziness in the count itself, since all the presences in geodesics are counted irrespective of the shape of the fuzzy shortest path distances. In order to retain such information in the model, the following measure of betweenness is proposed:

$$\widehat{b_k} = \sum_i \sum_j \frac{\widehat{g_{ijk}}}{\widehat{g_{ij}}} \; i \neq j \neq k \qquad (7)$$

The $\widehat{g_{ij}}$ represents a fuzzy count of the number of fuzzy geodesics. The computation of that count extends the crisp counterpart in the following way.

$$\widehat{g_{ij}} = (g_{ij}, \alpha(\widehat{d_{ij}}), \beta(\widehat{d_{ij}})) \qquad (8)$$

where α and β are the parameters from the output of the FSPP algorithm used. This way, the uncertainty in the triangular number is retained in the betweeenness measure.

The $\widehat{g_{ijk}}$ in the above formula can be modeled as follows.

$$\widehat{g_{ikj}} = (g_{ikj}, \qquad (9)$$

$$\alpha(pred(fsp_{ij}^k)) + \alpha(succ(fsp_{ij}^k)), \qquad (10)$$

$$\beta(pred(fsp_{ij}^k)) + \beta(succ(fsp_{ij}^k))) \qquad (11)$$

where $\alpha(pred(fsp_{ij}^k))$ is the α parameter of the direct fuzzy distance of the predecessor of actor k in the (fuzzy) geodesic from i to j. The rest of the elements can be interpreted similarly, with *succ* standing for the successor in the geodesic. Note that the expression entails directionality but the direction does not affect the final result.

3.2 Interpreting fuzzy betweeenness

The above measure for fuzzy betweenness may be interpreted as an extension of the original idea in two directions. On the one hand, it considers the parameters α and β in the fuzzy geodesics $\widehat{d_{ij}}$. Since betweenness is measuring the intermediation of an actor, it seems reasonable that these parameters are related to the parameters of the shortest distance. Since FSPP algorithms return fuzzy numbers with aggregated parameters, these can be directly translated to the measure. However, these parameters in general grow with the (crisp)length of the path so that it might be reasonable to find a transformation that compensates the effect of that length if the interest in betweenness is that of measuring distant sub-networks [1].

Regarding the count of times k is in between i and j, the interpretation varies. In this case, the main value of betweenness is finding actors filling structural holes. Then, the important aspect of their value is the degree of strength to the *adjacent* actors. The situation becomes more complicated if the actor bridges the subnetworks with more than one tie of similar intensity, however we will not consider that case here. In the case considered, the idea is that the parameters α and β are related to the uncertainty associated to the direct ties to k's adjacent nodes. This is clearly a driver of the information transfer [11] potential of the actors covering structural holes. Then, betweenness is a more effective indicator in that it reflects the uncertainty of ties. This could be combined with intensity to give a more informed measure, but this will entail moving away from the original idea of (crisp) betweenness.

4 Conclusions and Future Research

On-line social systems represent a case of social networks where the social contacts of a given actor are represented as non-valued graphs that reflect the structure of ties formed by chained invitations to connect. However, these systems provide a model of ties that does not consider the strength or intensity of the ties. This is in contrast with existing studies that deal with friendship dynamics and similar relationships, considering the intensity as a key element in the studies.

Modeling the intensity of ego-centered relations can be done by a representation of subjective intensities in the form of fuzzy numbers. Then, fuzzy arithmetics and extensions of typical social net measures can be used to provide better informed models for diverse practical applications. This paper has reported some initial possibilities for such kind of extended models and their application to basic functionality in on-line social network systems, including information and contact filtering, and finding relevant individuals. Concretely, measures of indegree centrality and betweenness have been explored and initial evidence on their appropriateness for users are reported.

Our intention is to extend and apply the metrics presented in this paper considering applications in two directions. First by implementing an extension to the current systems to accommodate models as those reported here can be realized thanks to the extensibility features provided by some of the social platforms such as `Facebook`. On the other hand, there is a need to explore additional, more complex models and test their properties for the concrete aspects of social relationships. This can be done by comparing the perceived appropriateness of the users comparing extended and standard versions for the same functionality.

References

[1] R. Burt. The network structure of social capital. *Research in Organizational Behavior*, 22(2):345–423, 2000.

[2] L. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

[3] M. Granovetter. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[4] F. Hernandes, M. T. Lamata, J. L. Verdegay, and A. Yamakami. The shortest path problem on networks with fuzzy parameters. *Fuzzy Sets and Systems*, 158(14):1561–1570, July 2007.

[5] Korfiatis. Social measures and flexible navigation on online contact networks. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, London, UK, 2007. IEEE.

[6] N. Korfiatis, M. A. Sicilia, C. Hess, K. Stein, and C. Shindler. Social network models for enhancing reference based web rankings. In D. H. Goh and S. Foo, editors, *Social information retrieval systems: Emerging technologies and applications for searching the web effectively.* IDEA Group Publishing, 2006.

[7] T.-S. Liou and M.-J. J. Wang. Ranking fuzzy numbers with integral value. *Fuzzy Sets Syst.*, 50(3):247–255, September 1992.

[8] P. Marsden. Network Data and Measurement. *Annual Review of Sociology*, 16(1):435–463, 1990.

[9] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.

[10] D. Robson. How to spot false friends on Facebook. *New Scientist*, 195(2613):28–28, 2007.

[11] F. Salvetti and S. Srinivasan. Local Flow Betweenness Centrality for Clustering Community Graphs. *Lecture notes in computer science*, pages 531–544, 2005.

[12] M. Selfhout, S. Branje, T. ter Bogt, and W. Meeus. The role of music preferences in early adolescents friendship formation and stability. *Journal of Adolescence*, 2007.

[13] K. tornblom and E. Fredholm. Attribution of Friendship: The Influence of the Nature and Comparability of Resources

Given and Received. *Social Psychology Quarterly*, 47(1):50–61, 1984.

[14] N. Way, K. Cowal, R. Gingold, K. Pahl, and N. Bissessar. Friendship patterns among African American, Asian American, and Latino adolescents from low-income families. *Journal of Social and Personal Relationships*, 18(1):29, 2001.

[15] J. Whitmeyer. A Deductive Approach to Friendship Networks. *The Journal of Mathematical Sociology*, 26(3):147–165, 2002.

[16] J. Whitmeyer and C. Yeingst. A dynamic model of friendly association networks. *Social Science Research*, 35(3):642–667, 2006.

[17] R. Yager. Ranking fuzzy subsets over the unit interval. *17th Symposium on Adaptive Processes,IEEE*, 17, 1978.