Measuring Betweenness Centrality in Social Internetworking Scenarios

Francesco Buccafurri*, Gianluca Lax, Serena Nicolazzo, Antonino Nocera, and Domenico Ursino

DIIES, University "Mediterranea" of Reggio Calabria
{bucca,lax,s.nicolazzo,a.nocera,ursino}@unirc.it

Abstract. The importance of the betweenness centrality measure in (on-line) social networks is well known, as well as its possible applications to various domains. However, the classical notion of betweenness centrality is not able to capture the centrality of nodes w.r.t. paths crossing different social networks. In other words, it is not able to detect those nodes of a multi-social-network scenario (called Social Internetworking Scenario) which play a central role in inter-social-network information flows. In this paper, we propose a new measure of betweenness centrality suitable for Social Internetworking Scenarios, also applicable to the case of different communities of the same social network. The new measure has been tested in a number of synthetic networks, highlighting the significance and effectiveness of our proposal.

1 Introduction

Centrality is one of the most important and widely used measures in Social Network Analysis [9,7,3]. A large variety of applications can benefit of this measure: for instance, think of information flow and strategic marketing, to name a few. Among the centrality measures, one of the most known is betweenness centrality. The betweenness centrality of a node is defined as the fraction of shortest paths between node pairs that pass through it. It is capable of measuring the influence of a node over the information spread through the network [1,10]. Due to its relevance in network analysis, this measure has been largely investigated, and several extensions, tailored to particular contexts, have been proposed in the past [11,5,6,2]. Unfortunately, the existing notion of betweenness centrality is not able to capture the centrality of nodes w.r.t. paths crossing different social networks (or different communities in the same social network). In other words, it is not able to detect those nodes of a multi-social-network scenario (called Social Internetworking Scenario) which play a central role in inter-social-network information flows. In this paper, we propose a new measure of betweenness centrality, called *cross betweenness centrality (CBC)*, suitable for Social Internetworking Scenarios, also applicable to the case of different communities of the same social network. Intuitively, the CBC of a node works by

* Corresponding author.

Y.T. Demey and H. Panetto (Eds.): OTM 2013 Workshops, LNCS 8186, pp. 666–673, 2013.

considering only paths which cross the different social networks about which we want to measure the interconnection capability of this node. We investigate the relationship between the new measure and the classical one also by analyzing some features of CBC. Finally, with the support of some synthetic networks, we substantiate the significance and the effectiveness of our measure also by showing where the classical one fails. The plan of this paper is as follows: in the next section, we formally define cross betweenness centrality. In Section 3, we present our tests. Finally, in Section 5, we draw our conclusions.

2 Cross Betweenness Centrality

Consider a Social Internetworking Scenario, from now on called SIS, consisting of a set $\mathcal{S} = \{S_1, \ldots, S_k\}$ of social networks. It can be represented as a graph $G = \langle N, E \rangle$, where N is the set of its nodes and E is the set of its edges. A node $a \in N$ belongs to exactly one social network of the SIS. We denote by $S(a)$ this social network. E is partitioned into two subsets E_f and E_m. E_f is said the set of *friendship edges* and E_m is the set of `me` *edges*. A `me` edge links the accounts of the same user in two different social networks and can be easily declared by the user himself [4]. E_f is such that for each $(a, b) \in E_f$, $S(a) = S(b)$, while E_m is such that for each $(a, b) \in E_m$, $S(a) \neq S(b)$.

An edge $(a, b) \in E_f$ means that the account b is a friend of the account a in the social network $S(a)$. An edge $(c, c') \in E_m$ means that the user of account c in $S(c)$ has declared a `me` edge between c and the account c' in $S(c')$. In this case, we say that c is a *bridge*. We observe that our definition of bridge does not correspond to the classical one defined in the context of single social networks, even though a strong relationship between the two notions can be argued.

Given a node $u \in N$, the classical definition of betweenness centrality is the following:

$$BC(u) = \sum_{s,t \in N, s \neq u, t \neq u} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from s to t, and $\sigma_{st}(u)$ is the number of those shortest paths passing through u.

This formula involves all the shortest paths. Thus, it is not able to measure how much the node u is central w.r.t. the information flow crossing different social networks. This could be a very important information. For instance, think of a music band and consider a SIS consisting of MySpace and LinkedIn. The capability of disseminating information from nodes of MySpace to nodes of LinkedIn is extremely relevant for the marketing goals of the band, which probably is already known in MySpace but unknown in LinkedIn. A bridge allowing the connection (and the subsequent information flow) between MySpace and LinkedIn can open a lot of new opportunities to the band. Therefore, the importance of this bridge in this SIS is extremely high even in case the paths involving it would be not numerous, which would imply a low value of its betweenness centrality. We thus need to revise the definition of betweenness centrality to our purpose, in order to bring out the importance of nodes w.r.t. the information flow crossing different social networks, as done for weak ties in classical studies on social networks [8].

Our definition of betweenness centrality in a SIS, which we call *cross betweenness centrality (CBC)*, is the following:

$$CBC(u, \Omega) = \sum_{s,t \in N, s \neq u, t \neq u, S(s) \neq S(t), S(t) \in \Omega} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

Here, Ω is any subset of $\mathcal{S}$.

This definition allows the computation of the cross betweenness centrality of a node to be limited (if this is desired) to a subset of the social networks of the SIS. Observe that:

$$BC(u) = CBC(u, \Omega) + CBC(u, \overline{\Omega}) + IBC(u)$$

where:

$$IBC(u) = \sum_{s,t \in N, s \neq u, t \neq u, S(s) = S(u) = S(t)} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

and

$$\overline{\Omega} = \mathcal{S} \setminus \Omega$$

We call $IBC(u)$ the *internal betweenness centrality* of u. Therefore, the betweenness centrality of a node consists of three components: The first two have an inter-social-network nature, whereas the third one is an intra-social-network component. Three important properties can be derived from the above equation:

1. $CBC(u, \Omega) \leq BC(u)$ for each $u \in N$ and for each $\Omega \subseteq \mathcal{S}$;
2. in the trivial case of a single-social-network SIS, $BC(u) = IBC(u)$;
3. if u belongs to a fragment of a social network not connected with the rest of the SIS, then $CBC(u, \Omega) = 0$ for each Ω.

Interestingly, CBC could be identically applied to the case of different communities of a single social network, by considering the social network nodes as partitioned into communities and by substituting the concept of social network in a SIS with the concept of community in a social network.

3 Tests

In this section, with the support of two synthetic networks, we show the significance and the effectiveness of our cross betweenness centrality measure.

The first network is shown in Figure 1. In this case we have a SIS consisting of three social networks whose nodes are colored in black, gray and white, respectively. We compute the betweenness centrality BC and the cross betweenness

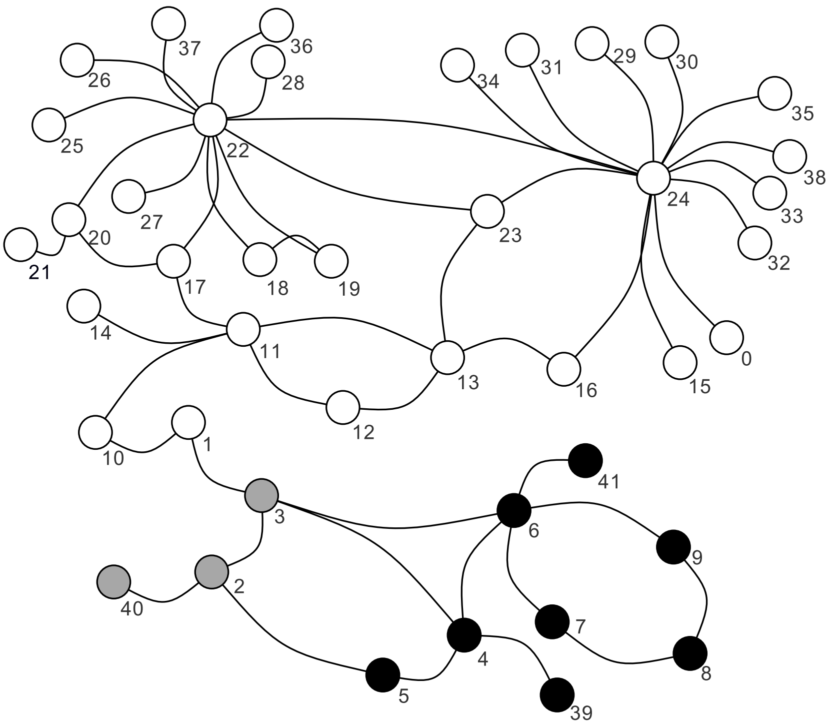

Fig. 1. A first synthetic network

centrality CBC for each node of the SIS. For CBC we choose Ω in such a way as to consider all the social networks of the SIS. In Table 1 we report the values of BC and CBC, along with the corresponding ranks, for the most significant nodes of the SIS.

From the analysis of Figure 1 and Table 1 we can observe that: *(i)* the three most ranked nodes for BC and CBC are totally different; *(ii)* the three most ranked nodes for BC are high-degree nodes and none of them is a bridge or is directly connected with a bridge; *(iii)* the three most ranked nodes for CBC are either bridges (e.g., nodes 1 and 3) or directly connected to bridges (e.g., node 10); *(iv)* none of the most ranked nodes for CBC is a high-degree node.

This example clearly shows how CBC is capable of capturing and representing a phenomenon high relevant for a SIS (since it is strictly connected with information spread), that BC is not able to capture in this scenario.

All these observations are further enforced by examining the second network, shown in Figure 2. In this case we have a SIS consisting of three social networks, where a great fragment of one of them is disconnected from the rest of the SIS. We compute BC and CBC as in the previous case. In Table 2 we report the corresponding values and ranks for the most significant nodes of the SIS.

Table 1. Values and ranks of BC and CBC for the most significant nodes of the SIS of Figure 1

Node	*BC*	*Rank*	*CBC*	*Rank*
11	802	1	614	4
22	771	2	374	6
24	734	3	220	8
1	660	5	660	1
3	644	6	644	2
10	696	4	636	3

Table 2. Values and ranks of BC and CBC for the most significant nodes of the SIS of Figure 2

Node	*BC*	*Rank*	*CBC*	*Rank*
3	264	1	0	7
10	120	2	0	7
24	102	3	0	7
21	66	5	50	1
17	48	6	48	2
16	42	7	42	3

All the observations drawn for the previous example are still valid, and further reinforced, in this case. For instance, the three most ranked nodes for BC belongs to the disconnected fragment of a social network of the SIS, and information passing through them has no chance to reach nodes of the other social networks. As a further observation, we evidence that this example confirms the third property of CBC derived at the end of Section 2.

4 Related Work

In this section, we survey the approaches proposed in the literature strictly related to the topic of betweenness centrality.

Betweenness centrality is capable of measuring the influence of a node over the information spread through the network. This concept has been widely investigated in [1,10]. In [1], the analysis of betweenness centrality of nodes in large complex networks. obtaining that it is increasing with connectivity as a power law with an exponent η, which is equal to 2 for trees or networks with a small loop density. In contrast, while a larger density of loops leads to $\eta < 2$. For scale-free networks, characterized by an exponent γ which describes the connectivity distribution decay, betweenness centrality is also distributed according to a power law with a non universal exponent δ.

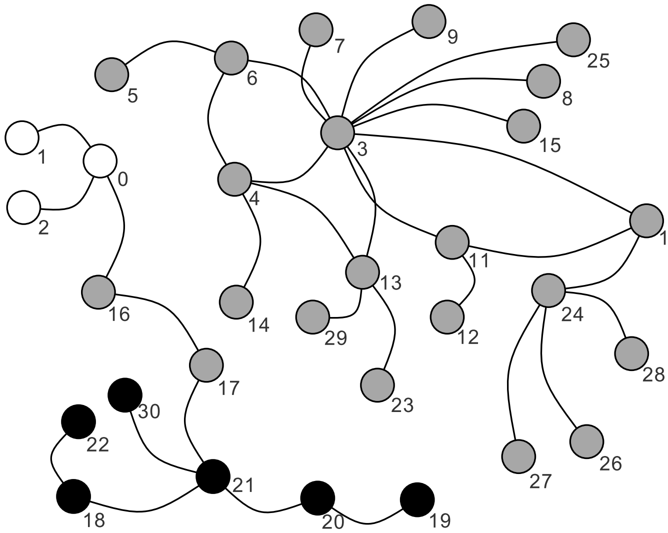

Fig. 2. A second synthetic network

In [10], Newman proposes a new betweenness centrality measure, called Random-Walk betweenness, that essentially counts all the paths between vertices (excluding those that do not actually lead from the designated source to the target), and which makes no assumptions of optimality. This measure is based on random walks between vertex pairs. It asks, in essence, how often a given vertex will fall on a random walk between another pair of vertices. This measure is particularly useful for finding vertices of high centrality that do not happen to lie on geodesic paths or on the paths formed by maximum-flow cut-sets. The author shows that the new betweenness centrality can be calculated using matrix inversion methods in a time that scales as the cube of the number of vertices on a sparse graph, making it computationally tractable for the networks typical of current sociological studies.

Due to its relevance in network analysis, several extensions of betweenness centrality, tailored to particular contexts, have been proposed in the past [11,5,6,2].

In [11], the authors generalize Freeman's geodesic centrality measures for betweenness centrality on undirected graphs to the directed case. This is an important step in the development of scientific propositions concerning social networks.

[5] extends the standard network centrality measures of degree centrality, closeness centrality and betweenness centrality to apply to groups and classes as well as individuals. The group centrality measures enable researchers to answer questions such as "how central is the engineering department in the informal influence network of this company?" or "among middle managers in a given organization, which are more central, the men or the women?" By means of these measures, they can also solve the inverse problem: "given the network of

ties among organization members, how can we form a team that is maximally central?" The authors define different group centralities and formalize a measure of group centrality efficiency, which indicates the extent to which a group's centrality is mainly due to a small subset of its members.

In [6], the authors look at the betweenness centrality of the ego in an ego network. They discuss the issues concerning normalization and develop an algorithm for computing the betweenness centrality score. The computation of all the ego betweenness scores for a whole network would be one order of magnitude faster than the computation of real betweenness centrality scores. The authors examine also the relationship between the betweenness centrality of an actor in her ego network and her betweenness centrality in the whole network. Even if, they show that there is no theoretical link between these two measures, they present a simulation study which indicates that the local ego betweenness centrality is highly correlated with the one of the actor in the complete network. Specifically, Ego betweenness centrality gives a good approximation of betweenness centrality in two situations, i.e. when all the actors have very similar betweenness centrality scores, and when there are highly differentiated scores.

In [2], the authors study the problem of analyzing multidimensional networks. They start from the consideration that complex networks have been receiving an increasing attention by scientific community, thanks also to the increasing availability of real-world network data. The aim of [2] is then to give the basis for multidimensional network analysis. The authors present a repertoire of basic concepts and analytical measures, which take the general structure of multidimensional networks into account. They test the framework on different real-world multidimensional networks, showing that the measures introduced are able to extract important and non-random information about complex phenomena in such networks.

All the notions of centrality described above have been conceived for single social network. Therefore, they are not able to capture the centrality of nodes w.r.t. paths crossing different social networks (or different communities in the same social network). This is the main contribution of our paper, which is the first one proposing a centrality measure specifically conceived for a multi-social-network scenario.

5 Conclusion

In this paper, we have presented cross betweenness centrality, a measure that adapts betweenness centrality to a SIS. In particular, we have defined this measure, we have analyzed its relationship with classical betweenness centrality, we have investigated some of its features, and we have highlighted its significance by means of synthetic networks also showing that the classical measure of betweenness centrality is not able to capture the centrality of nodes w.r.t. paths crossing different social networks. As pointed out in the introduction, this short paper presents our preliminary results about this issue. In the future, we plan to extend our research efforts in several directions. Specifically, we plan to conduct an experimental campaign to better evaluate the new measure in real SIS contexts, instead of in synthetic networks, also considering its application to different communities of a single social network. Finally, we plan to investigate the relationship between our notion of centrality and the concept of weak ties introduced in [8], on the basis of the similarity observed in Section 2.

Acknowledgment. This work has been partially supported by the TENACE PRIN Project (n. 20103P34XC) funded by the Italian Ministry of Education, University and Research.

References

1. Barthelemy, M.: Betweenness centrality in large complex networks. The European Physical Journal B-Condensed Matter and Complex Systems 38(2), 163–168 (2004)
2. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Multidimensional networks: foundations of structural analysis. In: World Wide Web, pp. 1–27 (2012)
3. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2(1), 113–120 (1972)
4. Buccafurri, F., Foti, V.D., Lax, G., Nocera, A., Ursino, D.: Bridge Analysis in a Social Internetworking Scenario. In: Information Sciences, pp. 224:1–224:18. Elsevier (2013)
5. Everett, M.G., Borgatti, S.P.: The centrality of groups and classes. The Journal of Mathematical Sociology 23(3), 181–201 (1999)
6. Everett, M.G., Borgatti, S.P.: Ego network betweenness. Social Networks 27(1), 31–38 (2005)
7. Freeman, L.C.: Centrality in social networks conceptual clarification. Social Networks 1(3), 215–239 (1979)
8. Granovetter, M.S.: The strength of weak ties. American Journal of Sociology 78(6), 1360–1380 (1973)
9. Katz, L.: A new status index derived from sociometric analysis. Psychometrika 18(1), 39–43 (1953)
10. Newman, M.E.J.: A measure of betweenness centrality based on random walks. Social Networks 27(1), 39–54 (2005)
11. White, D.R., Borgatti, S.P.: Betweenness centrality measures for directed graphs. Social Networks 16(4), 335–346 (1994)