Driving Global Team Formation in Social Networks to Obtain Diversity

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

DIIES, University of Reggio Calabria
Via Graziella, Località Feo di Vito
89122 Reggio Calabria, Italy
{bucca,lax,s.nicolazzo,a.nocera,ursino}@unirc.it

Abstract. In this paper, we present a preliminary idea for a crowdsourcing application aimed at driving the process of global team formation to obtain diversity in the team. Indeed, it is well known that diversity is one of the key factors of collective intelligence in crowdsourcing. The idea is based on the identification of suitable nodes in social networks, which can profitably play the role of generators of diversity in the team formation process. This paper presents a first step towards the concrete definition of the above application consisting in the identification of an effective measure that can be used to select the most promising nodes w.r.t. the above feature.

1 Introduction and Description of the Idea

It is well known that diversity is one of the key factors of collective intelligence in crowdsourcing [20]. On the other hand, it is clear that this concept fully confirms the famous principle summarized as *the strength of weak ties*, stated in the field of social networks [14]. The two worlds, social networks and crowdsourcing, have a strong overlap, as social-network users form a huge crowd. But, social-network crowd includes something more than the simple Web crowd. It has a friendship-based structure, embeds contents, and is full of knowledge about people. This opens a lot of opportunities that can reinforce the power of crowdsourcing (e.g., see [16]). For instance, consider the problem of dynamic formation of globally distributed teams for enterprisers [21]. Driving team formation can result in tangible benefits for the success of the team work, as a number of features of the individuals, such as expertise, should be considered. Social networks are repositories of a large amount of information about people, in which we can find the aimed features. But this is not enough. Indeed, it is not what a crowdsourcing process, thus spontaneous and evolutionary, requires. As a matter of fact, individuals in a social networks are not monads. So, not only friendship relationships allow the autonomous flooding of the network, enabling crowd formation, but the type of ties on which the crowd formation propagates can be dramatically important for the final quality of the global team: We have to hope that the most of crossed ties are weak, to fully reach the goal of diversity. Therefore, the

S. Casteleyn, G. Rossi, and M. Winckler (Eds.): ICWE 2014, LNCS 8541, pp. 410–419, 2014.

basic principle we are stating here is that global team formation in crowdsourcing, even though mostly spontaneous, should be driven in some way by taking advantage of social networks with the aim of maximizing diversity.

In this paper, we try to answer a simple question: How to translate the above principle in a social-web application aimed at improving the quality of global team formation in crowdsourcing? This is a preliminary research, as it contains an idea on how to do this, but it is focused on a very important aspect which is the starting point for transforming this idea into a concrete application. The idea is that team formation should be driven by weak ties, as mentioned earlier. But, how to find weak ties? From social network analysis we know that the concept of weak tie is related to behavioral aspects (for example, the number of interactions in the dyad), which are very difficult to capture in a feasible way. Actually, the concept of weak tie is also related to structural properties, since weak ties are typically bridges between two different communities. In particular, it is related to centrality measures [11]. Among these, betweenness centrality has a primary role, as it is the fraction of shortest paths between node pairs that pass through a node, and thus it is capable of measuring the influence of this node over the information spread through the network [1,17]. Therefore, to detect nodes that interconnect different communities we have to find nodes that are central in terms of betweenness centrality. In general, this search could appear difficult, if no restriction of the domain is done. The idea we present here is based on the consideration that some special nodes exist that exhibit explicitly a role of connectors between two different worlds. To understand this, we have to move towards the perspective of the social internetworking scenario (SIS) [18,3,7,6,8], in which the scope of the action of both people and applications is not confined to a single social network. As a matter of fact, in the current scenario characterized by hundreds of online social networks, a single user can join more of them. This leads to have membership overlap among social networks as expression of different traits of users' personality (sometimes almost different identities), also enabling, as side effect, the passage of information from one social network to another. As a consequence, membership overlap can be viewed as a feature that gives a specific power to users in terms of capability of connecting different worlds. But the good news is that, differently from a generic central node in a social network, a node interconnecting two different social networks (that we call *i-bridge*) often exhibits some explicit information showing this feature, thus it can be easily (automatically) identified. Indeed, often, a user shows in the home page of her account in a social network the link to her account in another social network. As a consequence, it is possible to define crawling strategies that privilege these nodes, allowing their easy discovery. One of these strategies is that called Bridge-Driven Search (BDS) presented in [9].

On the basis of this reasoning, our idea is to use BDS to find a number of seeds for our team formation, possibly iterating the process until a suitable stage is reached. But, the basic question is now: Are all i-bridges good for our purpose? How to compare different i-bridges in terms of capability of generating diversity? On the basis of the above considerations, one could think that the classic measure of betweennes centrality could be used to do this. In this paper, we show that the above hypothesis is not correct. In particular, we show that we need a measure of centrality that, differently from classic betweenness centrality, is able to take into account paths crossing distinct social networks in a different way from internal paths. Thus, we study a new measure of betweenness centrality, called *cross betweenness centrality (CBC)*, which allows us to characterize nodes in a social internetworking scenario in terms of importance w.r.t. inter-social-network information flows and to characterize candidate nodes in terms of suitability to be actors in global team formation.

It is clear that all the above reasoning appears well-founded only if the basic (even intuitive) assertion really holds that i-bridges play the role of ties connecting two social networks in a weak way. Thus, in this paper, as a first contribution, we prove the above claim through an experimental validation.

The plan of this paper is as follows. In the next section, we model our reference scenario. A description of the testbed adopted for the experimental campaign is described in Section 3. In Section 4, we prove the basic claim that i-bridges play the role of ties connecting two social networks in a weak way. In Section 5, we introduce the new measure, called *cross betweenness centrality (CBC)*, needed for our application to detect good i-bridges. An experimental proof about the need and the validity of CBC is presented in Section 6. In Section 7, a review of the related literature is discussed. Finally, in Section 8, we draw our conclusions and identify how to continue our research.

2 The Multi-Social-Network Model

In this section, we model our reference scenario, which is called *social internetworking scenario* (SIS) and takes into account the multiplicity of social networks and their interconnections through `me` edges. A *SIS* is a directed graph $G = \langle N, E \rangle$, where N is the set of *nodes* representing (social network) user accounts and E is the set of *edges* (i.e., ordered pairs of nodes) representing relationships between user accounts. Given a node $a \in N$, we denote by $S(a)$ the social network which a belongs to. 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.* E_f is such that for each $(a, b) \in E_f$, $S(a) = S(b)$, while E_m is such that for each $(a, b) \in E_m$, $S(a) \neq S(b)$. Each node a of G is associated with the (social network) account of a user joining the social network $S(a)$. An edge $(a, b) \in E_f$ means that the user having the account b is a friend of the user having the account a (both a and b belong to the same social network). An edge $(c, c') \in E_m$ means that c and c' are accounts of the same user in two different social networks. As a consequence, c is an i-bridge and a `me` edge interconnects the two social networks $S(c)$ and $S(c')$. Given a SIS $G = \langle N, E \rangle$, we call *the social networks of* G the set of social networks $\mathcal{S}$ such that for each $Q \in \mathcal{S}$ there exists $n \in N$ such that $S(n) = Q$. Observe that, the graph of a SIS differs from those underlying a single social network because of the presence of `me` edges, which connect nodes of different social networks. From now on, consider given a SIS $G = \langle N, E \rangle$.

3 Tools and Data for Our Experiments

Our experiments have been carried out on real-life data obtained by crawling online social networks. As for the crawling strategy, we used the well-known BFS [22], which performs a *Breadth First Search* on a local neighborhood of a seed it starts from. The crawling task was performed by means of the system $SNAKE$ [5], which is able to extract not only connections among the accounts of different users in the same social network, but also connections among the accounts of the same user in different social networks. These connections are represented by two standards encoding human relationships, namely XFN (XHTML Friends Network) and FOAF (Friend-Of-A-Friend). This way, we got a dataset consisting of five social networks, namely Flickr, LiveJournal, Google+, MySpace, and Twitter. They are compliant with the XFN or FOAF standards and have been largely analyzed in social network analysis in the past. Starting from this real-life dataset, we extracted several subgraphs for our tests. Each subgraph was obtained by randomly choosing an i-bridge node b and selecting all the nodes having a minimum distance from b less than or equal to 4 (observe that, due to the small diameter of real-life social networks, the chosen distance is significative). Figure 1 shows one of the subgraphs of the real-life dataset, in which black and white nodes stand for users belonging to Flickr and LiveJournal, respectively.

4 Are `me` edges Weak ties?

One of the most fascinating results of social network analysis is due to one of the fathers of this discipline and regards the concept of weak ties [14]. Although a complete notion of the strength of a tie can be given only if we consider dynamic and behavioral information, even the sole structural knowledge about a social network allows us to identify those ties that, informally speaking, connect two dense components keeping a low connection degree between them. The theory presented in [14] and confirmed by years of study on social networks, gives a strong importance to such ties, as they connect different communities, so they can be a formidable vehicle of cross contamination between them. The claim underlying our crowdsourcing application is that i-bridges are good candidates to be actors in a global team formation. But to prove this claim, the first step is to face this issue: Are `me` edges weak ties, in general?

In this section, we try to give an answer to the above question by conducting a suitable experiment on real-life social networks. To detect weak ties, we adopt the strategy proposed in [10]. In particular, we consider an edge e between the nodes n_1 and n_2 and we check if the removal of this edge would increase the distance between n_1 and n_2 to a value strictly more than two. If this occurs, then e can be considered as a weak tie. This experiment is carried out on the whole dataset described in Section 3 consisting of a set of $171,982$ normal edges (i.e., non-`me` edges) and a set M of 79 `me` edges. At the end of the experiment, the approach described above detected a set W of 5619 weak ties among the set E of $172,061$

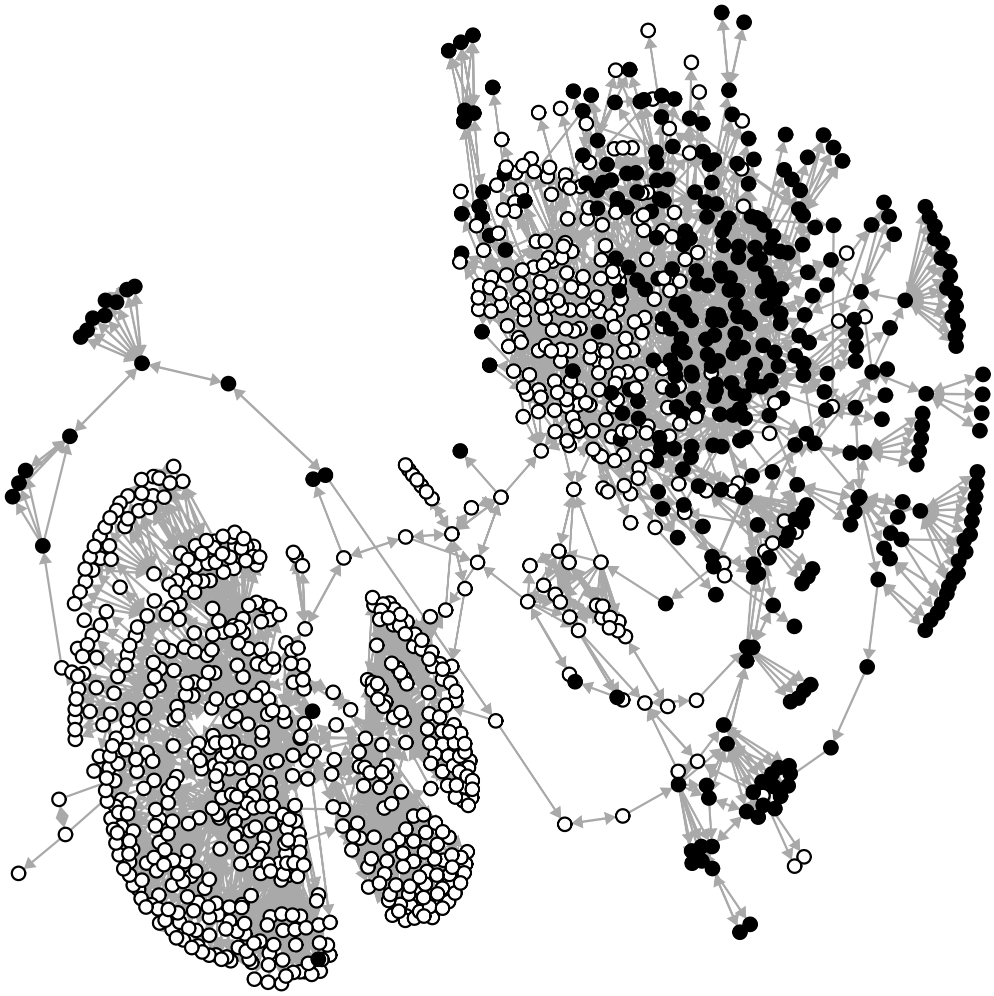

Fig. 1. Visualization of one subgraph

edges and 60 of them were also `me` edges. We calculated the percentage of `me` edges which are weak ties as $\frac{|M \cap W|}{|M|} = 0.76$, the percentage of `me` edges $\frac{|M|}{|E|} = 4.6 \cdot 10^{-4}$ and the percentage of weak ties $\frac{|W|}{|E|} = 3.3 \cdot 10^{-2}$. From the above results, we conclude that the probability that a `me` edge is a weak tie is high, whereas the probability that a generic edge is a weak tie is very low. Because $|M|$ and $|W|$ are much lesser than $|E|$, the result that $\frac{|M \cap W|}{|M|} = 0.76$ demonstrates that a strong relation between weak ties and `me` edges exists. Thus, the experiment concludes that the correlation between weak ties and `me` edges exists, so that we can sight potential powerful roles for such edges in a social internetworking scenario. It is worth remarking that the interpretation of `me` edges as weak ties means that, given a user u, her account in a social network S sees her account in another social network T (in case a `me` edge from S to T is established by u) as a weak tie. This means that u can be used as powerful disseminator of information across different communities, each belonging to a different social network. Obviously, the more the number of `me` edges of a user u, the higher her strength (in the

Granovetter sense [14]) in the network. Observe that the above conclusion is not in contradiction with the results given in [8], where the presence of `me` edges into user accounts has been proven to be assortative. Indeed, one could think that, if the friends of a user u assortatively declare `me` edges from the social network S to the social network T, as done by u, we obtain a very dense clique of users invalidating the result about the correspondence between `me` edges and weak ties. However, [8] shows that the *strict* (i.e., towards the same social network T) assortativity does not hold for the most representative real-life social networks (i.e., Facebook). In other words, the friends of u assortatively declare `me` edges from S to any other social network. Thus, the contradiction does not exist.

5 Measuring the Suitability of Diversity Generators: Cross Betweenness Centrality

After having proved, in Section 4, that `me` edges are weak ties, we know that our application is well-founded in the sense that all the i-bridges that we are able to find by using a crawling strategy as BDS [9] are good candidates to play the role of diversity generators in team formation. In this section we face a second important issue: Are all the candidates the same in terms of suitability to our application? In general, betweenness centrality (BC) [11] is used to detect weak ties (and also their *structural* strength). Unfortunately, as we will show in Section 6, the above claim cannot be applied to the case of i-bridges, in the sense that it is not able to measure their structural strength in terms of connectors of two social networks. Therefore, in this section, we introduce a new measure called *cross betweenness centrality (CBC)* to *rate* candidates in our application, overcoming the limits of BC. The need and the validity of CBC is shown in the next section. Recall that we are interested in a measure able to take into account paths crossing distinct social networks in a different way from internal paths. Even though the definition of betweenness centrality does not explicitly take into account the presence of the multiplicity of social networks, it could happen that the real-life structure of the interconnections among distinct social networks (i.e., i-bridges) is such that BC automatically favors nodes belonging to the frontier of each social network, as paths are in some way forced to cross them. Intuitively, the above claim is true if the density of the involved social networks is comparable. Otherwise, we expect that the most dense social network works as an accumulation point, biasing the centrality towards it. However, also in this case, the role of i-bridges is still crucial, so we would like not to miss it.

The definition of *cross betweenness centrality (CBC)* is the following. Let $\Omega \subseteq \mathcal{S}$. Given a node $n \in N$, we denote the *cross betweenness centrality* of n w.r.t. Ω as:

$$CBC(n, \Omega) = \begin{cases} \sum_{s,t \in N, s \neq n, t \neq n, S(s) \neq S(t), S(t) \in \Omega} \frac{\sigma_{st}(n)}{\sigma_{st}} & \mathit{if}\ \sigma_{st} > 0 \\ 0 & \mathit{otherwise} \end{cases}$$

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

Table 1. Results obtained for the first subgraph

Type of node	*BC*	*CBC*
Bridges	1,242,081.10	42.67
Power Users	1,543,513.59	4.43
Normal Users	3,795.34	0.01

Ω is a subset of the social networks of the SIS (see Section 2) and 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. In the definition of CBC, considered paths are only those (*i*) linking two nodes belonging to different social networks and (*ii*) having the target node (t) belonging to one social network in Ω (it does not matter whether the source node s belongs to a social network in Ω). In particular, we compute how many times the node n is involved in this kind of path. Interestingly, if n belongs to a fragment of a social network not connected with the rest of the SIS, then $CBC(n, \Omega) = 0$.

Observe that the following relation between cross betweenness centrality and the classical betweenness centrality can be proved: $BC(n) = CBC(n, \Omega) + CBC(n, \overline{\Omega}) + IBC(n)$, where $IBC(n) = \sum_{s,t \in N, s \neq n, t \neq n, S(s)=S(t)} \frac{\sigma_{st}(n)}{\sigma_{st}}$ and $\overline{\Omega} = \mathcal{S} \setminus \Omega$. A direct consequence of this results is that, in the trivial case of a single-social-network SIS, $BC(n) = IBC(n)$. Indeed, no inter-social-network contribution occurs.

6 Need and Validity of CBC: An Experimental Proof

In this experimental section, we show that BC is not able to capture the capability of i-bridges to be central w.r.t. cross-social-network paths. Then, we show that CBC is a measure that can be used in our application to compare different i-bridges in terms of suitability to be actors in global team formation. For this purpose, we partitioned nodes into three categories:

1. i-bridges, which are nodes with a `me` edge;
2. power users, which are non-i-bridge nodes whose degree is equal to, or higher than, the average degree of all nodes;
3. normal users, which are neither i-bridges nor power users.

Then, we computed the average values of BC and CBC for each category. The experiments presented in this section are carried out on two of the subgraphs described in Section 3. The results obtained for the first and second subgraphs are reported in Tables 1 and 2.

From the analysis of these tables, we note that there is no correlation between BC and node cathegory. Indeed, in Table 1, i-bridges and power users have comparable values, whereas normal users have a value that is about 3 magnitude orders lesser than the previous ones. By contrast, in Table 2, power users and

Table 2. Results obtained for the second subgraph

Type of node	*BC*	*CBC*
Bridges	105.01	43.33
Power Users	1,480,655.07	5.14
Normal Users	1,092,715.52	0.04

normal users have comparable values, whereas i-bridges have a value 4 magnitude orders lesser than the previous ones. Thus, it is evident that betweenness centrality is not able to correctly identify i-bridge nodes.

Consider now cross betweenness centrality. By looking at Tables 1 and 2, we observe that, for each node category, there is a great uniformity in the corresponding values. Even more interesting, i-bridges have a value of CBC always higher than power users (about one magnitude order), which, in turn, show a value much higher than normal users (about two magnitude orders). Therefore, the distinction among the three categories of nodes is evident by taking cross betweenness centrality into account. Thus, even our expectation about cross betweenness centrality is confirmed by analyzing real-life social networks.

7 Related Work

The concept of centrality, as applied to the context of human communication, was first introduced by Bavelas in 1948 [2]. He mainly focused on communication in small groups and he hypothesized a relationship between structural centrality and influence in group processes. More recently, Leavitt [15], Shaw [19] and Goldberg [13] proposed studies on speed, activity and efficiency in solving problems, on personal satisfaction and on leadership in small group settings. The concept of centrality is motivated by the idea that a person who is close to others can have access to more information, a higher status, more power, a greater prestige, or a greater influence [12] than others. Indeed, this person can facilitate or inhibit the communication of others and is, therefore, in a position to mediate their information access, power, prestige, or influence. Among all the centrality measures, betweenness centrality is one of the most popular, and its computation is the core component of a range of algorithms and applications. Both Bavelas [2] and Shaw [19] suggested that, when a person is strategically located in the middle of communication paths linking other users, she is central. A person in such a position can influence the group by holding or distorting information. By the way, the development of betweenness centrality is generally attributed to the sociologist Linton Freeman [11]. Over the past few years, betweenness centrality has become a popular strategy to measure node influence in complex networks, such as social networks. For this purpose, a lot of new metrics based on betweenness centrality have been already defined [17]. A concept strongly related to edge importance is edge classification. This task is usually performed on the basis of the kind (and, hence, of the "strength") of the relationship the edge represents. Under this assumption, an edge could be a strong or a weak tie. The concept of tie strength was introduced by Mark Granovetter in his very popular paper entitled "The Strength of Weak Ties" [14]. He identified four main features contributing to outline the strength of a tie, namely: amount of time, intimacy, intensity and reciprocal services. Finally, a first characterization of the nodes of a SIS has been proposed in [4]. However, no experimental validation has been provided therein.

8 Conclusion and Future Work

In this paper, we have presented a preliminary idea of a crowdsourcing application aimed at driving the process of global team formation to obtain diversity in the team. This paper presents a first step towards the concrete definition of the above application consisting in the identification of an effective measure that can be used to select seed nodes in the team formation. A first preliminary experimental validation has been provided showing that our idea is well-founded. The next steps are to further validate the new measure and to design the social web application in detail. We plan to do this in our future research.

Acknowledgments. This work has been partially supported by the TENACE PRIN Project (n. 20103P34XC) funded by the Italian Ministry of Education, University and Research and by the Program "Programma Operativo Nazionale Ricerca e Competitività" 2007-2013, Distretto Tecnologico CyberSecurity 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. Bavelas, A.: A Mathematical Model for Small Group Structures. Human Organization 7(3), 16–30 (1948)
3. Buccafurri, F., Foti, V., Lax, G., Nocera, A., Ursino, D.: Bridge Analysis in a Social Internetworking Scenario. Information Sciences 224, 1–18 (2013)
4. Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A., Ursino, D.: Measuring Betweenness Centrality in Social Internetworking Scenarios. In: Demey, Y.T., Panetto, H. (eds.) OTM 2013 Workshops 2013. LNCS, vol. 8186, pp. 666–673. Springer, Heidelberg (2013)
5. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: A system for extracting structural information from Social Network accounts. Technical Report. Available from the authors
6. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Crawling Social Internetworking Systems. In: Proc. of the International Conference on Advances in Social Analysis and Mining (ASONAM 2012), Istanbul, Turkey, pp. 505–509. IEEE Computer Society (2012)
7. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Discovering Links among Social Networks. In: Flach, P.A., De Bie, T., Cristianini, N. (eds.) ECML PKDD 2012, Part II. LNCS, vol. 7524, pp. 467–482. Springer, Heidelberg (2012)

8. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Internetworking assortativity in Facebook. In: Proc. of the International Conference on Social Computing and its Applications (SCA 2013), Karlsruhe, Germany, pp. 335–341. IEEE Computer Society (2013)
9. Buccafurri, F., Lax, G., Nocera, A., Ursino, D.: Moving from social networks to social internetworking scenarios: The crawling perspective. Information Sciences 256, 126–137 (2014)
10. Easley, D., Kleinberg, J.: Networks, crowds, and markets, vol. 8. Cambridge University Press, Cambridge (2010)
11. Freeman, L.C.: Centrality in Social Networks Conceptual and Clarification. Social Networks 1(3), 215–239 (1979)
12. Friedkin, N.E.: Theoretical foundations for centrality measures. American Journal of Sociology 96(6), 1478–1504 (1991)
13. Goldberg, S.C.: Influence and leadership as a function of group structure. Journal of Abnormal and Social Psychology 51(1), 119–122 (1955)
14. Granovetter, M.: The strength of weak ties. American Journal of Sociology 78(6), 1360–1380 (1973)
15. Leavitt, H.J.: Some effects of a certain communication patterns on group performance. Journal of Abnormal and Social Psychology 46(1), 38–50 (1951)
16. Lim, S., Ncube, C.: Social networks and crowdsourcing for stakeholder analysis in system of systems projects. In: Proc. of the International Conference on System of Systems Engineering (SoSE 2013), Maui, Hawaii, USA, pp. 13–18. IEEE (2013)
17. Newman, M.: A measure of betweenness centrality based on random walks. Social Networks 27(1), 39–54 (2005)
18. Okada, Y., Masui, K., Kadobayashi, Y.: Proposal of Social Internetworking. In: Shimojo, S., Ichii, S., Ling, T.-W., Song, K.-H. (eds.) HSI 2005. LNCS, vol. 3597, pp. 114–124. Springer, Heidelberg (2005)
19. Shaw, M.E.: Group structure and the behavior of individuals in small groups. Journal of Psychology 38(1), 139–149 (1954)
20. Surowiecki, J., Silverman, M.: The Wisdom of Crowds. American Journal of Physics 75(2), 190–192 (2007)
21. Vukovic, M.: Crowdsourcing for enterprises. In: Proc. of the International Conference on Services-I 2009, Los Angeles, CA, USA, 2009, pp. 686–692. IEEE Computer Society Press (2009)
22. Ye, S., Lang, J., Wu, F.: Crawling online social graphs. In: Proc. of the International Asia-Pacific Web Conference (APWeb 2010), Busan, Korea, pp. 236–242. IEEE (2010)