Community-Based Measures for Social Capital

Christopher Spratt[1], Jun Hong[2], Kevin McAreavey[1] and Weiru Liu[1]

[1] University of Bristol, Bristol BS8 1TH, UK,
{chris.spratt, kevin.mcareavey, weiru.liu}@bristol.ac.uk
[2] University of the West of England, Bristol BS16 1QY, UK,
jun.hong@uwe.ac.uk

Abstract. Social capital is the value that can be derived from connections between individuals in a social network. The most common forms are bonding and bridging social capital, resulting from connections with similar and diverse groups of individuals respectively. In this paper we propose a novel community-based model for measuring bonding and bridging social capital in a social network. Some previous measures of bonding and bridging capital depend on node attributes, which are often difficult to obtain. Other measures overcome this limitation by relying purely on network structure but are limited to direct connections only for bonding capital and indirect connections only for bridging capital. Our structural measures for bonding and bridging capital are independent of attributes and account for both direct and indirect connections. We experimentally validate our measures on a collaboration network extracted from DBLP, and our results show a strong correlation with standard measures of academic success.

Keywords: social capital, community detection, social network analysis

1 Introduction

At its simplest, social capital is the value that can be placed on the connections between individuals [1], providing insight into the ways in which both societies and individuals prosper [2].

The two most widely acknowledged and implemented forms of social capital are the duals of bonding and bridging capital [3]. Putnam [4] described bonding as being social connections amongst similar groups of individuals and bridging as social connections with diverse groups of individuals. Some have focused on diversity and similarity in terms of the structure of the graph [5], others in terms of the characteristics of individuals from hobbies to their role in a workplace [4]. Those who previously developed algorithms to measure social capital in graphs continued this focus on bonding and bridging. Subbian et al. [6] took a distance based approach, where direct connections were considered as bonding capital and indirect connections as bridging capital. The series of works by Smith et al. [7,8] and Sharma et al. [9] focused on characteristics, for example Smith et al. making use of blog.

An issue with taking a characteristic based approach is the difficulty involved in extracting those characteristics of an individual that are not directly available. Approaches such as analysing the content that an individual generates have the potential to be computationally intensive as well as being domain-dependent. Furthermore while Subbian et al. overcome these restrictions in their work, they solely differentiate between individuals based on the distance between each other in the network.

Communities have been recognised as being important structural components in social networks that impart valuable information about the structure and function of the networks [10]. This in addition with our intuition that individuals cluster together on the basis of some set of characteristics, would indicate that community membership is a useful way of characterising individuals given that membership could then be as a proxy for these core characteristics.

We propose a methodology that makes use of the community membership of individuals as the basis for similarity and diversity to avoid the need for attributes. Our approach is therefore to make the conceptualisation that:

- Bonding capital is resultant from connections between those who are members of the same community.
- Bridging capital results from connections between those who are members of disparate communities.

From this conceptualisation we propose a new model of social capital and develop a multi-stage algorithm to calculate the bonding and bridging capital for the individuals in a social network. We then go on to extensively test on a academic collaboration network extracted from DBLP data. Our contributions can be summarised as:

1. The development of an attribute-independent social capital model.
2. The development of structural measures for bonding and bridging capital that account for both direct and indirect connections.
3. Evaluation against various measures of academic success at both individual and group levels.

The rest of the paper is structured as follows: Section 2 describes the related work. Section 3 presents our proposed social capital model. We describe our experimental validation in Section 4. Finally Section 5 concludes with proposals for future work.

2 Related Work

One of the challenges in capturing social capital in a computational model is the variety of conceptualisations proposed over the years. From Putnam's popularisation of the concept [1], the focus has been on differentiating between bonding social capital and bridging social capital.

Within computer science there have only been two core sets of works attempting to extract social capital from social networks, firstly those by Smith et al. [7] then followed by Subbian et al. [6]. Their approaches can be separated by their interpretation of the similarity and diversity that bonding and bridging social capital is based upon respectively.

Smith et al. [7] whose focus was initially on the determination of social capital in blogging networks, took a characteristic based approach to bonding and bridging capital. Their approach combines the use of an explicit affinity network containing the explicit relationships between people and an implicit affinity network (IAN), mapping the underlying characteristic based affinities between people.

Subbian et al. [6] took a very different structural interpretation of bonding and bridging capital. They proposed a value allocation approach, which calculates the social capital value contained within the network based upon the sum of the benefits derived from the shortest paths between all pairs of individuals in the network, with the benefits decaying exponentially to path length. This results in a measure of the density of connections throughout the network. Social capital value is allocated to individuals in the network on the basis of the frequency that they can be found on the shortest paths between individuals relative to the lengths of the paths, representing their fractional contribution. The bonding capital of an individual in the network is captured by the benefits resultant from their immediate neighbours. On the other hand the bridging capital of an individual is captured by benefits resulting from their non-immediate neighbours, as well as in the allocation function. Though Subbian et al. [6] account for bonding and bridging capital, the output of their algorithm is a combined measure of social capital assigned through their value allocation.

Smith et al are restricted in their model by the difficulty of extracting characteristic information. The advantage of their approach over Subbian et al's work is how it makes explicit the importance of similarity and diversity. Our work overcomes the limitations of both works with explicit conceptualisation of similarity and diversity and being independent of extracted characteristics.

3 Community-based Social Capital Measures

In this section, we formulate the problem of measuring an individual's bonding and bridging social capital in a social network. First, we define a graph model of social networks. Second, we define communities in a social network. Third, we define community-based measures for similarity and diversity between individuals in a social network. Finally, we define measures for bonding and bridging social capital based on similarity and diversity measures.

3.1 Social Network Model

An undirected graph is a tuple $G = (V,E)$, where V is a set of nodes and E is a set of undirected edges $E = \{\{n_i,n_j\} \mid n_i,n_j \in V\}$. A social network is represented as an undirected graph $G = (V,E)$, where each node $n \in V$ represents an individual and each edge $\{n_i,n_j\} \in E$ represents a social tie between nodes n_i and n_j in the social network hence $n_i \neq n_j$.

Given an undirected graph, $G = (V,E)$, a path, p, between two nodes $n \in V$ and $n' \in V$ is defined as a sequence of nodes as follows:

1. If there exists an edge between n_1 and n_2, i.e. $\{n_1,n_2\} \in E$, then $p = (n_1,n_2)$ is a path, and p is said to be of length 1.
2. If $p = (n_1,n_2,...,n_m)$ is a path, and there exists an edge between n_m and n_{m+1}, i.e., $\{n_m,n_{m+1}\} \in E$ and $n_{m+1} \neq n_i$ for $i = 1,...,m-1$, then $p' = (n_1,n_2,...,n_m,n_{m+1})$ is a path, p' is said to be of length m.

The set of edges in path $p = (n_1, ..., n_{m+1})$ is defined as $E(p) = \{\{n_i, n_i + 1\} \mid i = 1, ..., m\}$. Let P denote the set of paths in $G = (V, E)$, we use $P(n, n') \subseteq P$ to denote the set of paths from node n to node n'. The path length between nodes n and n', $PathLength(n, n')$, is defined as the length of the shortest path in $P(n, n')$.

3.2 Community Detection

The concept of communities is intuitive as society offers a wide variety of possible forms of communities: families, groups of friends on social networks, colleagues in the same department, villages, cities, sports clubs, online forums, etc. A community is traditionally thought of as a tightly-knit group of nodes with more connections amongst its members than between its members and the other part of the network (Girvan and Newman, 2002).

In this paper we take a probabilistic view of communities in social networks. Instead of comparing the densities of edges inside and outside communities in a social network, we focus on the probability that nodes share edges within a community. The existence of communities implies that nodes interact more strongly with the other members of their community than they do with the members of the other communities. This is the reason why edge densities end up being higher within communities than between them. We can formulate this modern view in a probabilistic way; a node of a community has a higher probability to form edges with the other nodes of the community than with the nodes in the other communities.

Definition 1 (Community). *Let $G = (V, E)$ be a graph representing a social network. We say $V_1 \in V$ is a community in G if and only if for any node $n_i \in V_1$, $p(n_j \in V_1 | \{n_i, n_j\} \in E) > p(n_j \notin V_1 | \{n_i, n_j\} \in E)$, where $p(n_j \in V_1 | \{n_i, n_j\} \in E)$ and $p(n_j \notin V_1 | \{n_i, n_j\} \in E)$ represent the probabilities that $n_j \in V_1$ and $n_j \notin V_1$ respectively, given that there exists an edge $\{n_i, n_j\} \in E$.*

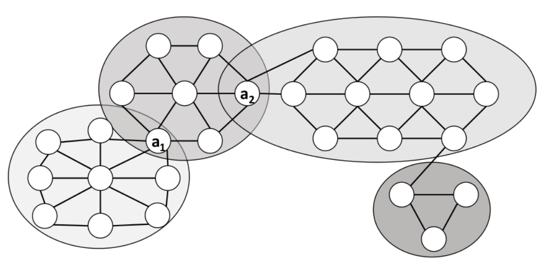

Fig. 1: Communities detected in a graph.

In this paper, we are particularly interested in how individuals interact inside a community and across communities. In real life situations an individual could be a member of multiple communities. For example, in a workplace an individual may be involved in multiple friendship groups formed around various different interests that they have. As shown in Figure 1, it is possible for individuals to be members of multiple communities up to the total number of communities in the graph as with nodes a_1 and a_2.

3.3 Similarity and Diversity

Communities are groups of individuals that share some common set of characteristics. Our intuition for similarity is that the more communities two individuals share, the more characteristics of those communities they share and the more similar they are. The inverse would hold true for diversity, the more disjoint communities they have, the more of their characteristics are disjoint.

Definition 2 (Similarity). *Let $G = (V,E)$ be a social network with A_i the set of communities that have node $n_i \in V$ as a member and A_j the set of communities that have node $n_j \in V$ as a member. Similarity is a function $Sim : V \times V \rightarrow [0,1]$ defined as:*

$$Sim(n_i,n_j) = \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \quad (1)$$

In this definition, the Jaccard coefficient [11] captures the number of communities n_i and n_j share relative to the total number of communities that n_i and n_j are members of. The similarity between n_i and n_j ranges from 0 to 1, where 0 signifies that n_i and n_j share no communities hence completely dissimilar, while 1 shows that they are members of the exact same communities hence completely similar.

Example 1. As shown in Fig. 2, given that $C = \{V_1,V_2,V_3\}$ is a set of communities such that $V_1 = \{n_1,n_5\}$, $V_2 = \{n_1,n_2\}$, and $V_3 = \{n_3,n_4\}$, we have: $Sim(n_1,n_5) = 0.5$, $Sim(n_3,n_4) = 1.0$, and $Sim(n_2,n_4) = 0.0$.

Definition 3 (Diversity). *Let $G = (V,E)$ be a social network with A_i the set of communities that have node $n_i \in V$ as a member and and A_j the set of communities that have node $n_j \in V$ as a member. Diversity is a function $Div : V \times V \rightarrow [0,1]$ defined as:*

$$Div(n_i,n_j) = \frac{|(A_i \cup A_j) \setminus (A_i \cap A_j)|}{|A_i \cup A_j|} = 1 - Sim(n_i,n_j) \quad (2)$$

In this definition, the Jaccard distance [11] captures the proportion of non-overlapping communities out of the total number of communities between n_i and n_j. The diversity between n_i and n_j ranges between 0 and 1, where 0 signifies that n_i and n_j are completely similar, while where 1 shows that they are completely diverse.

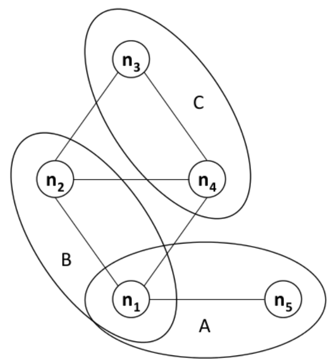

Fig. 2: Example graph with 3 communities

Example 2. As shown in Fig. 2, given that $C = \{A, B, C\}$ is a set of communities such that $A = \{n_1, n_5\}$, $B = \{n_1, n_2\}$, and $C = \{n_3, n_4\}$, we have: $Div(n_1, n_5) = 0.5$, $Div(n_3, n_4) = 0.0$, and $Div(n_2, n_4) = 1.0$.

3.4 Bonding and Bridging Social Capital

Social capital can take many forms [12]. In this paper, we focus on bonding and bridging social capital due to their prominence and conceptual accessibility. There have been many different definitions of these concepts but core to the majority is that bonding capital is resultant from the connections between a group of similar individuals and bridging between those who are in diverse groups. This has been popularised by Putnam [4] who described bonding as social capital that "enforces exclusive identities and homogeneous groups" where bridging "encompass[es] people across diverse social cleavages".

We define connectivity between a pair of nodes in a social network as a measure of the impact each of them has on the other on a social capital measure. We now define one possible interpretation of such impact as a function of the distance between the pair of nodes.

Definition 4 (Connectivity). *Let $G = (V, E)$ be a social network. Connectivity is a function $Con : V \times V \to \mathbb{R}^+$ defined as follows:*

$$Con(n_i, n_j) = e^{-\lambda PathLength(n_i, n_j)} \tag{3}$$

where $\lambda \in [0, 1]$ be a decay constant and $PathLength(n_i, n_j)$ is the length of the shortest path between n_i and n_j.

Intuitively, the connectivity between a pair of nodes in a social network decays exponentially to the distance between the pair in the network. How quickly the connectivity decays over distance is controlled by the decay constant, λ.

Example 3. As shown in Fig. 1, with $\lambda = 0.1$, the connectivity between nodes n_5 and n_3 can be calculated as such: $PathLength(n_5, n_3) = 3$, $Con(n_5, n_3) = e^{-0.1 \times 3}$, $Con(n_5, n3) = 0.74$

We define bonding capital as a social capital measure of how well an individual in a social network is connected to a group of similar individuals in the network.

Definition 5 (Bonding Capital). *Let $G = (V,E)$ be a social network. Bonding social capital is a function $\Lambda : V \to \mathbb{R}^+$ defined as:*

$$\Lambda(n_i) = \sum_{n_j \in V \setminus \{n_i\}} Sim(n_i,n_j) \times Con(n_i,n_j) \quad (4)$$

The bonding capital of an individual, n_i, in a social network is calculated as the sum of similarities between n_i and every other individual, n_j, in the network where the similarity between the pair, $Sim(n_i,n_j)$, is weighted by the connectivity between the pair, $Con(n_i,n_j)$.

Example 4. As shown in Fig. 2, with λ of 0.1, the Bonding Capital for node n_5 can be calculated as follows:
$\Lambda(n_5) = Sim(n_5,n_1) \times Con(n_5,n_1) + Sim(n_5,n_2) \times Con(n_5,n_2) + Sim(n_5,n_3) \times Con(n_5,n_3) + Sim(n_5,n_4) \times Con(n_5,n_4) = 0.5 \times 0.9 + 0 \times 0.81 + 0 * 0.74 + 0 \times 0.74 = 0.45$

We define bridging capital as a social capital measure of how well an individual in a social network is connected to diverse groups of individuals in the network.

Definition 6 (Bridging Capital). *Let $G = (V,E)$ be a social network. Bridging social capital is a function $M : V \to \mathbb{R}^+$ defined as:*

$$M(n_i) = \sum_{n_j \in V \setminus \{n_i\}} Div(n_i,n_j) \times Con(n_i,n_j) \quad (5)$$

The bridging capital of an individual in a social network is calculated as the sum of diversities between n_i and every other individual, n_j, in the network where the diversity between the pair, $Div(n_i,n_j)$, is weighted by the connectivity between the pair, $Con(n_i,n_j)$.

Example 5. As shown in Fig. 2, with $\lambda = 0.1$, the Bridging Capital for node n_5 can be calculated as follows:
$M(n_5) = Div(n_5,n_1) \times Con(n_5,n_1) + Div(n_5,n_2) \times Con(n_5,n_2) + Div(n_5,n_3) \times Con(n_5,n_3) + Div(n_5,n_4) \times Con(n_5,n_4) = 0.5 \times 0.9 + 1 \times 0.81 + 1 \times 0.74 + 1 \times 0.74 = 2.74$

4 Experiments

In the absence of ground truth of social capital in an existing social network, we indirectly evaluated our model through the benefits resulted from bonding and bridging social capital. We used a dataset created from the DBLP database with additional citation data from ACM [13]. The dataset contains approximately 2.2M papers published between 1936 and 2013.

We extracted a collaboration network from the dataset, in which each node represents an author of one of the papers from the dataset and each edge represents a co-authorship between two nodes if they have co-authored at least one paper together. The network contains roughly 1.4M undirected edges and 400K nodes. For every author we calculated the total number of citations from all of their papers, and the total number of citations from all of their papers published at each of the publication venues.

From this collaboration network we took two random 10K samples of nodes, and calculated their bonding and bridging social capital values. We tried to establish the relationships between social capital values of these nodes and the measures of their success in the dataset, including the total number of citations and the H-index for each node in the dataset.

Given social capital is viewed as as both an individual and collective property, we also examined our social capital measures at the group level. In the context of the DBLP dataset the most readily available groupings of individuals are the publication venues of the papers in the dataset. Based on the total number of citations, and H-Index of each paper published at a publication venue, we created two similar measures of success for each publication venue, including the total number of citations and H-index for each venue by aggregating the number of citations for each paper published at the venue, and then calculated the H-Index for each venue. For the two sample sets, there are approximately 2,500 publication venues.

Using the same 10K random samples we calculated the total bonding and bridging capital scores for each publication venue in the dataset by aggregating the bonding and bridging capital scores of each author published at the venue.

4.1 Implementation

For community detection we primarily made use of the BIGCLAM model as introduced by Leskovec and Yang [14]. BIGCLAM assumes that the probability of an edge between two nodes being generated in a social network is reflected in the commonality of their affiliations with communities, fitting the probabilistic view of communities. BIGCLAM views overlapping communities as being densely connected, as opposed to prior work on detecting overlapping communities which views overlapping communities as being sparsely connected.

A matrix of affiliation strengths between nodes and communities is fitted through an optimisation process that minimises the difference between the network that matrix would generate, and the observed network. The matrix is then transformed into a set of hard affiliations by only counting those nodes which exceed a strength threshold as members of a communities. The threshold is set such that if two nodes belong to a community then the probability that an edge forms between them is higher than the background edge probability. The background edge probability is a small probability that any two nodes that can be connected in a global community of all nodes ε.

We also examined how the model performed with communities detected by Blondel et al.'s [15] Louvain Community Detection Method as a benchmark. Louvain is both highly scalable to large networks and has been widely studied. It is a comparatively simple method which depends on the density of interaction between nodes to assign its communities. More specifically it optimises the modularity of the communities contained within a network, the number of edges between members of a community against the number of edges the members have with nodes outside the community. Though the technique does not allow for the detection of overlapping communities, it does not invalidate our definition of communities given that a non overlapping technique can be viewed as solving the same problem but at a lower level of granularity.

The implementation of our model was a simple breadth first exploration of the graph for each node, calculating the total bonding and bridging social capital based on pregenerated community assignments. To reduce the run-time of our social capital algorithm we placed a connectivity limit on the algorithm's exploration. This prevented the algorithm from exploring those nodes in the network that have a minimal impact on the node under examination.

4.2 Results

We performed experiments with a connectivity limit of 0.01 and $\lambda = 1$, BIGCLAM (implemented in C++ as part of the the SNAP package [3]) set to detect 4000 communities (on average one community per 100 nodes), and Louvain community detection as implemented in the Python-Louvain package[4].

We examined a number of hypotheses. At an individual level we examined whether the bonding or bridging social capital of a node had a positive correlation with their total number of citations and H-Index. At the group level we tested whether there was a strong correlation between the total bonding and bridging social capital of a publication venue, and the total number of citations and H-index of the venue. The results of these can be seen in Table 1 and 2 respectively, with the values being Pearson's correlation coefficients between the sets of data for the samples of nodes.

As can be seen in Table 1 we found that the correlation between our bonding and bridging measures and the number of citations an individual has received was relatively weak, where the correlation between the measures and an individual's H-Index was stronger. Consistently across all community detection techniques and the sample datasets we found that bridging had a higher correlation with the success measures than bonding. This would suggest that collaborations are good indicators of success, intercommunity collaborations can lead to greater prospects. Furthermore we found similar strength of correlation across the two random sample datasets. This indicates that we are capturing a wider trend in the dataset.

Interestingly we found that the Louvain based social capital measures consistently outperformed those using BIGCLAM detected communities across the samples. This suggests that the granularity of information on the structure of the network provided by the much larger number of communities detected through Louvain (circa 23,000) outweighs the information gained by a smaller, overlapping set of communities detected through BIGCLAM.

As shown in Table 2, we found that the correlations between success measures for publication venues and their aggregated social capital were consistently stronger.

[3] https://github.com/snap-stanford/snap
[4] https://python-louvain.readthedocs.io/en/latest/

	Total Citations /Bonding	Total Citations /Bridging	H Index /Bonding	H Index /Bridging
Sample 1 Louvain	0.217	0.344	0.307	0.571
Sample 2 Louvain	0.193	0.261	0.323	0.586
Sample 1 BIGCLAM	0.142	0.200	0.337	0.351
Sample 2 BIGCLAM	0.125	0.166	0.349	0.367

Table 1: Individual Level R-Values

	Total Citations /Bonding	Total Citations /Bridging	H Index /Bonding	H Index /Bridging
Sample 1 Louvain	0.644	0.697	0.669	0.737
Sample 2 Louvain	0.601	0.679	0.634	0.733
Sample 1 BIGCLAM	0.616	0.649	0.662	0.697
Sample 2 BIGCLAM	0.609	0.641	0.667	0.696

Table 2: Group Level R-Values

Once again the Louvain based measures had a higher degree of correlation to those produced using BIGCLAM. We also found a slightly stronger degree of correlation between bridging capital and the success measures than between bonding capital and the success measures at both individual and group levels. These results are not unexpected given the link between bridging social capital and advancement, and the link bonding social capital and support. The academic success measures we chose can be considered as being more directly connected as to the advancement of individuals.

5 Conclusions

In this paper we proposed a novel community based model for measuring bonding and bridging social capital in a social network. We developed community-based similarity and diversity measures, independent of extracted attributes that unlike prior works recognises the similarity and diversity between indirectly connected nodes in a community and cross communities in a social network respectively. These measures allowed us to have a closer interpretation of Putman's original definitions of an individual's bonding and bridging social capital in a social network. This interpretation recognises that the bonding social capital of an individual results from their social connections with a similar group of people in the same community, and the bridging social capital of an individual results from diverse groups of people outside the individual's own communities.

The correlations between our social capital measures with academic success at both individual and organisational levels indicate that we are indeed capturing social capital as we are able to detect its associated benefits. The trends in our experimental results also indicate that bridging capital yields better results than bonding capital in terms of academic success, which as mentioned in the previous section is not unexpected given bridging social capital's stronger links to advancement.

Our current model of social capital provides community-based measures for bonding and bridging social capital, which provides a good basis for a more elaborate social capital model when the characteristics of individuals in a social network are available. In such a case, characteristic based similarity and diversity can be combined with community-based similarity and diversity to create more elaborate similarity and diversity measures. A further direction we intend to explore is to use machine learning techniques to learn a predictive social capital model, which can predict an individual's social capital given a set of features capturing various structural properties of an individual in a social network.

Acknowledgements

This work received funding from the EU's Horizon 2020 research and innovation programme through the DEVELOP project, under grant agreement No. 688127.

References

1. Robert D Putnam. Bowling alone: America's declining social capital. *Journal of democracy*, 6(1):65–78, 1995.
2. Scott E Seibert, Maria L Kraimer, and Robert C Liden. A social capital theory of career success. *Academy of management journal*, 44(2):219–237, 2001.
3. Michael Woolcock. Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and society*, 27(2):151–208, 1998.
4. Robert D Putnam. *Bowling alone: The collapse and revival of American community*. Simon and Schuster, 2001.
5. Nan Lin. Building a network theory of social capital. *Connections*, 22(1):28–51, 1999.
6. Karthik Subbian, Dhruv Sharma, Zhen Wen, and Jaideep Srivastava. Finding influencers in networks using social capital. *Social Network Analysis and Mining*, 4(1):1–13, 2014.
7. Matthew Smith. *Implicit affinity networks*. PhD thesis, Brigham Young University, 2007.
8. Matthew Smith, Christophe Giraud-Carrier, and Samuel Stephens. Measuring and reasoning about social capital: A computational framework. *A Computational Framework for Social Capital in Online Communities*, page 27, 2011.
9. Rajesh Sharma, Kevin McAreavey, Jun Hong, and Faisal Ghaffar. Individual-level social capital in weighted and attributed social networks. In *Proceedings of the 9th International Workshop on Mining and Analyzing Social Networks for Decision Support (MSNDS'18)*, to appear.
10. Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. Community detection in social media. *Data Mining and Knowledge Discovery*, 24(3):515–554, 2012.
11. Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
12. Tom Schuller, Stephen Baron, and John Field. Social capital: a review and critique. *Social capital: Critical perspectives*, pages 1–38, 2000.

13. Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998. ACM, 2008.
14. Jaewon Yang and Jure Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.
15. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.