A Human Communication Network Model

Oksana Pichugina[1], Babak Farzad[2]

[1] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
pichugina_os@mail.ru
[2] Brock University, St. Catharines, Canada
bfarzad@brocku.ca

Abstract. A number of attributed formation models based on Erdos-Renyi and Barabasi-Albert random graph models are presented. One of them is a Human Communication Network (HCN) model based on time restrictions on face-to-face communication. Construction of this weighted network requires a few numerical parameters and allows to transform any unweighted node attributed network into weighted. This transformation helps solving numerous problems in Network Analysis such as community detection, network topology inference, etc. Understanding nature of human communication networks allows to solve many practical problems starting with fast spreading any information and innovation through the networks and ending with detecting key people, collaboration with whom helps achieving different goals.

Keywords. Social Networks, Community Detection, Attributed Networks, Node Partition, Random Graphs

Key Terms. Network Decoration, Community Detection, Random Graphs

1 Introduction

Network Analysis is an area of research that has been studied intensively lately. Researches investigate structural characteristics of different networks, network formation models, and many other related questions. Among a variety of networks, social networks, which reflect a diversity of people relationships, are a priority [5], [7]. Study of social networks is important since it helps understanding how our world is organized, what place each of us takes in it, how this situation affects us and how the knowledge can be used to achieve our goals. Social networks are characterized by heterogeneity of nodes and edges, sparsity, high average clustering coefficient, small average shortest path length and power-law degree distribution, existing observable and tightly bound groups of elements called communities [5]. Most of these properties are united in "small-world networks" and "scale-free networks" concepts [1],[4]. Many attempts have been made to construct social networks, but still no satisfactory solution to simulate all the listed properties is found [1],[4],[5]. It is also important to find efficient ways of these community detection (CD) [6],[8]. We believe that the key

ICTERI 2016, Kyiv, Ukraine, June 21-24, 2016

in qualitative CD in social networks is in using the heterogeneity (making them multi-layer ones) and study the issue of such networks formation.

2 Definitions and Notations

Definition 1. *[6] A social network is a hybrid graph, which is represented in the form:*

$$G = (V, E, \Lambda, \Lambda'), \tag{1}$$

where V is the set of nodes (the social network's users), E is the set of edges (these users relationships), Λ and Λ' contain an information about attributes related to each node $v \in V$ and each edge $\{u, v\} \in E$, respectively.

The network represented in the form (1) is an *attributed network* if $\Lambda \cup \Lambda' \neq \emptyset$. So, any attributed network representing individuals' relationships is social.
Let $\Lambda \in \mathbb{R}^{n \times K}$, $\Lambda' \in \mathbb{R}^{m \times K'}$ hence V, E are of size $|V| = n$, $|E| = m$. Initially, we consider an unweighted *node-attributed network* $G = \{V, E, \Lambda\}$, $K \geq 1$. Then we assign weights to its edges (*decorate* the edges) and come to consideration of a weighted network $G^w = \{V, E, \Lambda, \Lambda'\}$ with Λ' being a matrix-column of edge weights ($K' = 1$). G^w is a *node-edge-attributed* and used then for CD.
Introduce some notations: $G^{[\cdot]} = \{V, E^{[\cdot]}, \Lambda\}$ - is an unweighted node-attributed network with an adjacency matrix $B^{[\cdot]} = (b_{ij}^{[\cdot]}) \in \mathbb{R}^{n \times n}$. After decoration $E^{[\cdot]}$ by weights, the new weighted network is denoted by $G^{w[\cdot]} = \{V, E^{[\cdot]}, \Lambda, \Lambda'\}$ and its weighted adjacency matrix (WAM) - by $A^{[\cdot]} = (a_{ij}^{[\cdot]}) \in \mathbb{R}^{n \times n}$.
The *node degree* $d_i^{[\cdot]}$ of a node $v_i \in V$ in $G^{[\cdot]}$ is the number of its incident edges: $d_i^{[\cdot]} = |N_i^{[\cdot]}|$ where $N_i^{[\cdot]} = \{u \in V : u \leftrightarrow v_i, \{u, v_i\} \in E^{[\cdot]}\}$. The *node strength* $s_i^{[\cdot]}$ of $v_i \in V$ is a sum of weights of its incident edges in $G^{[\cdot]}$. In terms of adjacency and weighted adjacency matrices, these values are: $d_i^{[\cdot]} = \sum_j b_{ij}^{[\cdot]}$, $s_i^{[\cdot]} = \sum_j a_{ij}^{[\cdot]}$.
A network cover is a division of the network nodes $\mathcal{C} = \{C_l\}$ satisfying $\bigcup_{l=1}^{L} C_l = V$. If in the division the node clusters $C_l, l \in J_L = \{1, ..., L\}$, are pairwise disjoint, then it is called *a network partition*.
Assume that the nodes are decorated by K discrete attributes $\{AT^k\}$, $\Lambda = (at_i^k)$ where at_i^k is the value of AT^k for a node v_i, and there are L_k different values of AT^k. Let $AC_l^k \in V$ be a set of nodes with l-th value of AT^k. We call it a *node attribute cluster* (AC) and denote a G-partition into ACs related to different values of AT^k by $\mathcal{AC}^k = \{AC_l^k\}_{l \in J_{L_k}}$ ($n_l^k = |AC_l^k|$). Let $G^{[\cdot]k}$ be a $G^{[\cdot]}$-subnetwork related to AT^k. A *sum of unweighted networks* $\{G^k\}_k$ of the same node set V is an unweighted network $G = \{V, \bigcup_k E^k, \Lambda\}$. A *linear combination of weighted networks* $\{G^{wk}\}_k$ of a node set V is a weighted network $G^w = \{V, \bigcup_k E^k, \Lambda, \Lambda'\}$ with a WAM $A = \sum_k \alpha_k A^k$ where $\{\alpha_k\} \subset \mathbb{R}$ are coefficients of this linear combination. Let $\omega(G^{[\cdot]})$ be denoted a *weight* of a network $G^{[\cdot]}$ ($\omega(G^{[\cdot]}) = \sum_{ij} a_{ij}$). A network of a weight one is a *normalized network*.
The networks' linear combination is a *weighted network sum* if

$$G^w = \sum_k W^k \cdot G^{wk} \text{ where } \{W^k\} > \mathbf{0}, \ \sum_k W^k = 1. \tag{2}$$

3 Motivation

Let us consider a social network. Suppose that, in addition to basic information about the node and edge sets, there is available some extra information about the nodes and edges features (social semantic networks are highly helpful here [2]). These additional characteristics are called attributes and the procedure of their complementing is decoration of the network [3], [5] resulted in creation of an attributed network [8]. Applying CD on the network we, typically, get communities closely related to one node attribute and this dominant attribute does not allow us to observe communities in other layers related to the rest, less important, node attributes. For instance, in the humankind network the dominant attribute would be belongingness to families. If we are interested in study communities, say, in work place, then the family division is an obstacle on this way. However, if it is possible to transform the network into weighted, moreover, to assign edge weights to each layer subnetworks of the multi-layer network, the problem of multi-layer CD (MLCD) can be solved. For that we just detect the dominant attribute and extract the corresponding subnetwork from consideration repeating then the procedure on the remaining network.
The crucial part of the approach is constructing edge sets of the one-layer subnetworks and distributing weights within them. The first one is a problem of the attributed network formation considered in Sect.4.1 (the edge inference problem [5]), the second one is the edge attribute inference problem [5]. The last one we solve for a social network of people face-to-face communication in Sec.4.2.

4 Human Communication Network Models

At this section we touch formation of attributed networks. We are wondering how an attributed network (1) is formed if the information about nodes V and their attributes Λ is known. In other words, we review formation of an edge set E and its attributes Λ' and refer to them as Problems 1 and 2, respectively.

4.1 Attributed Network Formation

We consider a number of ways to solve Problem 1. For convenience, we interpret the presented network formation models in terms of communication of people spending a time together during common activities/interests (AIs). Here nodes are people and their AIs are the nodes' attributes.
Model 1 - an association network model. An association network G^a [5] is an example of an attributed network where links exist between any nodes with common attributes. It can be interpreted as a network of virtual contacts of people with common interests where supporting such contacts does not need anything.
The auxiliary network G^{wk} corresponds to each activity/interest (AI) AT^k; G^w is representable as a weighted network sum (2) of K networks, which are collections of complete graphs: $G^{wk} = \bigcup_{l \in J_{L_k}} K_{n_l^k}$. Thus the network G^a is a cover of

K overlapping V-partitions by a disjoint union of complete graphs.
Model 2 - an attributed networks model based on Erdos-Renyi Model. Suppose that for existing an edge a similarity of node attributes is necessary, but not enough because of randomness. Similar to Model 1, we represent the network G^w by (2). Edges in G^{wk} are created randomly with probability p_l^k between two nodes v_i, v_j sharing the l-th value of the attribute AT^k. Hence G^k is a node partition by Erdos-Renyi Random Graphs (ERRGs) [4]: $G^{wk} = \bigcup_{l \in J_{L_k}} ERRG(p_l^k, n_l^k)$ and the resulting network G^w is an overlapping of K partitions by ERRGs.
In terms of human communication, Model 2 simulates a real situation where a group of people is formed simultaneously. Contacts of each user occur randomly without analysing any prior information due to its inaccessibility. The communication can be established on a regular basis only if these people actually have common interests. Different type of contacts are formed independently.
Model 3 - an attributed networks model based on Barabasi-Albert Model. In comparison with Model 2, here we review a situation where a group of people is formed gradually. First of all, group members aspire to contacts with popular and authoritative colleagues in each area of expertise. First, these contacts are formed for the most important AIs, then for the less significant. A chance to clarify common interests is higher if the contact already exists.
As before, G^w is a weighted network sum (2). $\{G^{wk}\}$ are formed consecutively by k in accordance with decreasing priorities of node attributes. For each k an edge set E^k is formed between nodes with the same value of AT^k consecutively by i with probabilities depending on degrees of all preceding nodes $\{d_{i'}^k\}_{i'<i}$ and parameters p^k, p'^k ($p^k \leq p'^k$) for new and previously established contacts.
There are many ways of a generalisation to attributed networks of Barabasi-Albert Preferential Attachment Model [1]. For instance, each auxiliary network G^k is formed as follows: disjoint subsets of nodes of different ACs are connected by preferential attachment and then the isolated subnetworks are connected forming the whole node partition $\mathcal{AC}^k$. These all partitions are united into a cover with respect to node attributes priorities, Λ, and pre-assigned order of the nodes arising. The network layers are dependent regardless we consider the case $p^k = p'^k, \forall\, k$ (Model 3.1) or another one ($\exists k : p^k < p'^k$). Model 3.1 simulates a node partition by Barabasi-Albert Graphs (BAGs). Each G^{wk} can be represented in a manner of Models 1, 2: $G^{wk} = \bigcup_{l \in J_{L_k}} BAG(n_l^k, \alpha_l^k)$ where α_l^k is the power of preferential attachment in AC_l^k.

4.2 The Human Communication Model

The models presented in Sect. 4.1 - Model 2 and Model 3 - are able to simulate networks of real, face-to-face contacts implying requirements to spend time for keeping in touch. Suppose an edge set E was formed according to Models 2 or 3. To finish the network G^w formation, Problem 2 has to be solved and the matrix Λ' be formed. Here we present a way to distribute edge weights according to assumptions typical, in our opinion, for real people interaction. We will refer to the obtained network as a Human Communication Network (HCN).

The HCN model assumptions. *Condition 1.* People AIs had already formed; *Condition 2.* Connections between people are possible if they have common AIs; *Condition 3.* Each person distributes uniformly the time t^k allotted for supporting a contact related to the AI AT^k between friends of this interest;
Condition 4. For everyone possibility of the communication is restricted by time T. If for a person the time is not enough for supporting his/her contacts, then the time allotted for supporting a contact related to the AT^k and $AT^{k'}$ is distributed proportionally to t^k and $t^{k'}$, respectively;
Condition 5. If two persons with the same interest are ready to devote time to each other, then, if necessary, they come to a compromise following certain rules. Formalise Conditions 1-5 in terms of the WAM A. We rewrite (2) in the form:

$$G^w = \sum_k G^{w'k}, \text{ where } G^{w'k} = W^k \cdot G^{wk}. \tag{3}$$

In addition to G^w satisfying Conditions 1-5, we introduce networks G^{w*}, $G^{w'*}$ satisfying Conditions 1-3 and 1-4, respectively.
Similarly to the (3), G^{w*} and $G^{w'*}$ are representable as networks sums: $G^{w*} = \sum_k G^{*k}$, $G^{w'*} = \sum_k G'^{*k}$ where G^{*k}, G'^{*k} are subnetworks of G^{w*} and $G^{w'*}$ related to AT^k. Respectively, the following holds for the corresponding WAMs:

$$A = \sum_k A'^k, A^* = \sum_k A^{*k}, A'^* = \sum_k A'^{*k}. \tag{4}$$

Let a set of v_i, v_j common attribute values be found as follows: $\mathcal{E}_{ij} = \{k : at_i^k = at_j^k\} \subseteq J_K$. Then $N_i^{[\cdot]k} = \{v_j \in N_i^{[\cdot]} : k \in \mathcal{E}_{ij}\}$ is a set of v_i-neighbours with the same AT^k-value as v_i in $G^{[\cdot]}$. We expand the notations of the node degree and strength from the set $N_i^{[\cdot]}$ into the sets $\{N_i^{[\cdot]k}\}_k$: a) $d_i^{[\cdot]k} = |N_i^{[\cdot]k}|$ is the node attribute AT^k-degree of $v_i \in V$ in $G^{[\cdot]}$; b) $s_i^{[\cdot]k} = \sum_{v_j \in N_i^{[\cdot]k}} a_{ij}^{[\cdot]k}$ - is the AT^k-strength of v_i in $G^{[\cdot]}$. Respectively, the node strength in $G^{[\cdot]}$ is $s_i^{[\cdot]} = \sum_k s_i^{[\cdot]k}$.

1. We start with assigning edge weights in G^{w*}:
   (a) Condition 1 says that the network G^{w*} is decorated by discrete attributes AT^k and the matrix Λ is known;
   (b) Condition 2 means that the links are formed only by similarity of the node's attributes, hence if $i, j : \mathcal{E}_{ij} = \emptyset \Rightarrow \{v_i, v_j\} \notin E$.
   (c) Condition 3 allows to determine the ratio of A^{*k}-elements: if i, j, j', k, k' such that $at_i^k = at_j^k$, $at_i^{k'} = at_{j'}^{k'}$, then

$$\frac{at_{ij}^{*k}}{a_{ij'}^{*k'}} = \frac{t^k}{t^{k'}}. \tag{5}$$

   Since there is no restrictions on the communication time in G^{w*}, it implies that all of the contacts are supported at the appropriate level. So, the weights in G^{w*}, G^{w*k} can be assigned with respect to the maximal needed time t^k:

$$\forall i, j : k \in \mathcal{E}_{ij} \; a_{ij}^{*k} = t^k; \; a_{ij}^* = \sum_{k \in \mathcal{E}_{ij}} t^k. \tag{6}$$

Notice that A^* is symmetric thus G^{w*} is undirected. The communication time of each person depends on the number of the contacts of each type therefore the node strengths in G^{w*k}, G^{w*} are defined as follows: $s_i^{*k} = d_i^k t^k$, $s_i^* = \sum_{k,j} a_{ij}^{*k} = \sum_k s_i^{*k} = \sum_k d_i^k t^k$. In terms of the HCN model, the values s_i^*, s_i^{*k} can be interpreted as the time that a person i could devote for the communication overall and for the particular AI, correspondingly.

2. Moving on to the network $G^{w'*}$, we add Condition 4 - the time restriction - to the network G^{w*}. This condition determines how much time a person i is ready to spend for supporting each AI-contact depending on his/her priorities and the number of these contacts. It can be expressed as the restriction on node strengths by T-value: $\forall i\ s_i'^* \leq T$. If it holds, then the above restriction holds and $G^{w'*} = G^{w*}$, otherwise the weights a_{ij}^{*k} are scaled to meet the time restriction:
$$a_{ij}^{'*k} = \nu_i^* a_{ij}^{*k} \tag{7}$$
where the scaling parameter ν_i^* depends on the node i strength: $\nu_i^* = min\left(1, \frac{T}{s_i^*}\right)$. Substitution (7) into (5) yields: $\frac{a_{ij}^{'*k}}{a_{ij'}^{'*k'}} = \frac{t^k \cdot \nu_i^*}{t^{k'} \cdot \nu_i^*} = \frac{t^k}{t^{k'}}$. It means that each person distributes his/her own time independently from each other and guided common priorities $\overline{W}$ accumulated in $\bar{t} = (t^k)$: $\overline{W} = \bar{t}/|t|$. Find the WAM of $G^{w'*}$ by (4),(7):
$$a_{ij}^{'*} = \sum_k a_{ij}^{'*k} = \nu_i^* \sum_k a_{ij}^{*k} = \nu_i^* \cdot a_{ij}^*. \tag{8}$$
The weights $a_{ij}'^*$, $a_{ji}'^*$ determine how much time a person i is ready to devote for communication with a person j and vice versa. It is clear that normally these are different values, $a_{ij}'^* \neq a_{ji}'^*$. Thus the network $G^{w'*}$ is directed that does not display a face-to-face communication.

3. To describe the real situation, we consider constructing the final network G^w from $G^{w'*}$. By adding Condition 5, the abstract directed network $G^{w'*}$ is transformed into the undirected G^w with weights equal to time that both persons - i and j - actually devote to each other. The weights are obtained as a result of a compromise between these persons who are ready to spend together not the same time.
Let persons i and j have a real contact ($\mathcal{E}_{ij} \neq \{\emptyset\}$) and are looking for a compromise ($a_{ij}'^* \neq a_{ji}'^*$). The result of their common decision can be expressed as function of these weights $a_{ij} = f(a_{ij}'^*, a_{ji}'^*)$. The function $f(.)$ can be chosen in different way. For instance, we choose a simple averaging: $a_{ij} = \frac{1}{2}(a_{ij}'^* + a_{ji}'^*)$. Then, by (8) and due to a symmetry of A^*, we have:
$$a_{ij} = 0.5(\nu_i^* a_{ij}^* + \nu_j^* a_{ji}^*) = 0.5 \cdot a_{ij}^*(\nu_i^* + \nu_j^*). \tag{9}$$
Distribution of weights within $\{G'^k\}$ is obtained from (4), (6), (9): $a_{ij} = \sum_k a_{ij}^{'k} = \frac{\nu_i^* + \nu_j^*}{2} \sum_k a_{ij}^{*k} = \frac{\nu_i^* + \nu_j^*}{2} \sum_{k \in \mathcal{E}_{ij}} t^k = \frac{\nu_i^* + \nu_j^*}{2} \sum_k t^k b_{ij}^k$ wherefrom
$$a_{ij}^{'k} = 0.5(\nu_i^* + \nu_j^*) t^k b_{ij}^k,\ a_{ij}^k = a_{ij}^{'k}/W^k = 0.5|\bar{t}|(\nu_i^* + \nu_j^*) b_{ij}^k. \tag{10}$$

4.3 Human Communication Network Simulation

Example 1 - Model 2 simulation. First, we demonstrate a solution of Problem 1 for Model 2 (see Sect. 4.1). Parameters of a simulated node-attributed network G are: the order $n = 60$, the number of node attributes $K = 3$, the nodes are divided randomly into $\{L_k\}_k = \{5, 4, 6\}$ attribute clusters of the same sizes: $(n_l^k) = (12^5, 15^4, 10^6)$. The result of the simulation with parameters $(p_l^k) = (0.3^5, 0.3^4, 0.5^6)$ is shown in Figure 1.

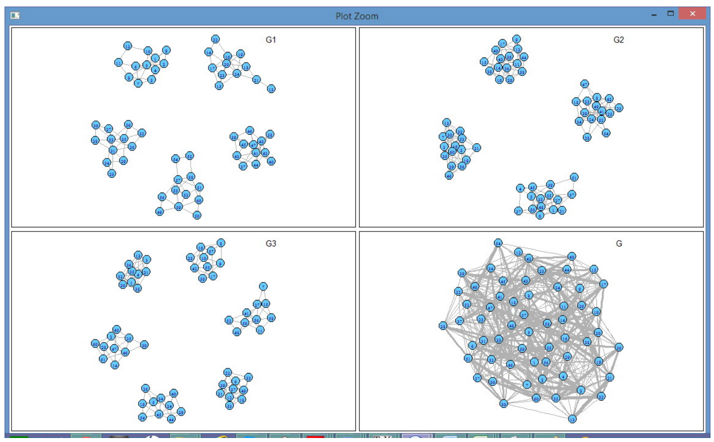

Fig. 1. Model 2 - the weighted network G^w and its subnetworks $G^1 - G^3$

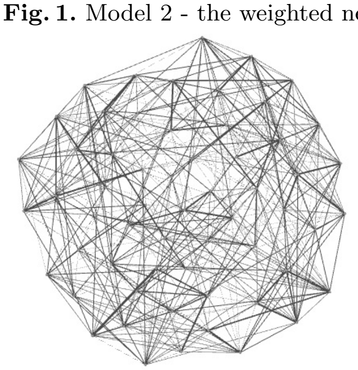

Fig. 2. The HCN G^{wI}

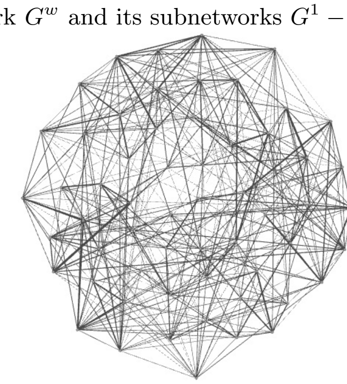

Fig. 3. The HCN G^{wII}

Example 2 - HCN Model 2 simulation. We took the unweighted network G from Example 1 and converted it into the HCN-Model 2 network (see Sect. 4.2) decorating edges by weights according to (9), (10). Two values of the time resource $\overline{T} = (T^I, T^{II})$ and the vector $\bar{t} = (t^1, t^2, t^3) = (4, 3, 2)$ are used. The vector of priorities of AIs is $\overline{W} = \frac{(4,3,2)}{|(4,3,2)|} = (0.45, 0.33, 0.22)$. We constructed two networks G^{wI}, G^{wII} corresponding to T^I, T^{II}. The time restrictions are

chosen in the following way: a) in the network G^{wI} for majority, 80%, of people the time T^I is sufficient to support their contacts completely; b) for the network G^{wII} the situation is opposite - most, 80%, of people should distribute their time resource T^{II}. For the simulated in Example 1 network these parameters are $\overline{T} = (56, 40)$. In Figures 2-3 we can see the resulted HCNs and observe that edge weights in G^{wI} are more heterogeneous than the ones in G^{wII}. Most likely, the reason is in absence in G^{wI}, in most cases, of necessity to redistribute the time resource. After normalizing G^w, the weights of the subnetworks $\{G^{wk}\}$ are $(\omega(G^{wk})) = (0.772, 1.330, 0.962)$, hence they are all not normalised and G^{w2} is the "haviest".
Results of community detection. CD on G does not show community structure in the network whilst CD on G^w quite accurately yields the partition $\mathcal{AC}^2$ into ACs related to AT^2, namely, in 80% of cases two ACs of $\mathcal{AC}^2$ were detected correct, rest two - with one error each in G^{wI}.

5 Conclusions and Future Work

The presented Human Communication Network (HCN) model demonstrates an approach to reconstructing missing network information about edges and edge weights based on node attributes and assumptions on nature of interaction in the networks. To the edge inference problem we apply an extension of Erdos-Renyi and Barabasi-Albert random graph models to multi-layer node attributed networks. There is shown that, in spite of interconnection of HCN layers, CD is running better in these networks decorated by weights.
The results we are planning to expand to other kinds of networks and use for designing new MLCD algorithms and solving node attribute inference problems.

References

1. Albert, R., Barabsi, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47-97 (2002)
2. Breslin, J., Passant, A., Decker, S.: The Social Semantic Web, 2010 edition. ed. Springer, Heidelberg; New York (2009)
3. Brning, J., Geiler, V.A., Lobanov, I.S.: Spectral Properties of Schrodinger Operators on Decorated Graphs. Mathematical Notes. 77, 858-861 (2005)
4. Erds, P., Rnyi, A: On random graphs, I. Publicationes Mathematicae (Debrecen). 6, 290-297 (1959)
5. Kolaczyk, E.D.: Statistical Analysis of Network Data, Springer Series in Statistics. Springer New York, New York (2009)
6. Lin, Z., Zheng, X., Xin, N., Chen, D.: CK-LPA: Efficient community detection algorithm based on label propagation with community kernel. Physica A: Statistical Mechanics and its Applications. 416, 386-399 (2014)
7. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, 1 edition. ed. Cambridge University Press, Cambridge; New York (1994)
8. Zhou, Y., Cheng, H., Yu, J.X.: Clustering Large Attributed Graphs: An Efficient Incremental Approach. In: 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 689-698 (2010)