Attributed Networks Formation Models with Application in Social Network Analysis

Oksana Pichugina

National Aerospace University "Kharkiv Aviation Institute", Chkalova Street 17, Kharkiv, 61070, Ukraine

Abstract
In the paper, networks characterized by vertices with a number of attributes (attributed networks) are studied. The networks cover a wide class of real-world ones, particularly social networks, making them especially attractive for further development. We analyze how the attributed networks are formed and present several mathematical models built based on their specifics and utilizing well-known classes of graphs – full ones, Erdös-Rényi and Barabasi-Albert random graphs. The computational experiment part includes simulation of test networks for all presented formation models, evaluating their metrics, and confirming their social networks properties. Barabasi-Albert graphs' based model best demonstrated these features. The results can be further used in Community Detection, Cluster Analysis, missing network attribute's restoration and other related Network Analysis problems.

Keywords 1
social networks, attributed networks, graph partition, graph cover, aggregation, Erdös-Rényi Random Graphs, Barabasi-Albert Model, complete graph.

1. Introduction

Network Analysis (NA) is a research field developed intensively in recent years. Researches in NA study different networks, particularly their social and structural characteristics, investigate statistic and dynamic behaviour of networks, design their formation models, single out their specific features, and solve many other related problems. Among all networks, social ones (SNs) that explore and reflect people and relationships between them are an absolute priority.

Why studying and deep understanding SNs is so important? First of all, it provides an understanding of how the world around us is organized. In turn, this allows us to realize what place we occupy in this global human network and how this understanding and knowledge can be used to achieve our goals. At the moment, many features of social networks have been derived. For example, it is known that in the networks there is a so-called "small-world" effect is observable. It manifests itself in a few handshakes between any two people on Earth. At the same time, despite the sparsity of SNs, dense subgraphs called communities are always present in the networks. Many researchers tried to design an ideal model of a social network. However, their attempts have not been crowned with success so far, although they managed to reproduce every single feature of SNs.

This paper is dedicated to modelling social networks and other ones in which vertices and edges are decorated with some discrete-valued attributes (attributed networks, ATNs). We propose three mathematical models for such networks based on the combination of isolated random graphs associated with different attributes and their values. Then we confirm experimentally that such networks are closer to SNs than the ones known so far.

2. Related Works and Motivations

Let us hypothetically imagine that we aim to study the global network of humanity and its community structure. Assume that we have full information about this network, about its participants, contacts and, most importantly, about the strength of these contacts. Thus, we are given full information about the weighted graph displaying this network. Also, imagine that we have access to the most powerful supercomputer in the world for conducting community detection (CD) for analysis in the global network. Evidently, the result of the CD will be quite predictable and will provide the division of people into families since relative relationships are the strongest.

The result of the CD is fully predictable; it would be a division by families because normally people spend most of the time with their families and because family ties are very strong. However, when we set up the experiment, we expected to get something we did not know before. We expect to see the network from different sides, for example, to see the relationship between friends or colleagues. However, in this global humankind network, many other communities are present. They combine people connected with other commonalities than family ties. In fact, setting the problem of the global community detection, we are generally interested in those hidden communities. We are interested in how to single out groups of colleagues in this network, groups of classmates, people with common interests etc. All these communities will be hidden in the network under the first layer of family relationships. Let us imagine that we managed to single out properly the weighted network corresponding to family relations. Then, having removed this dominant layer from the global network, we come to the next network where other layers of communities such as colleges', classmates' ones become visible and therefore detectable by community detection algorithms (CDAs). Having now analyzed the obtained network, repeated CD in it and associated the derived communities with a certain attribute, it becomes possible to single out a second dominant subnetwork. Now, by subtracting it from the graph, we can detect less significant communities hidden under the subnetwork. In such a way, hidden information in the global network becomes visible and detectable.

In NA terminologies, the human network vertices represent people, and edges reflect relationships between people. Each person is unique, and his uniqueness can be represented by a unique combination of quantitative and qualitative characteristics playing the role of the vertices attributes. The same works for the network's edges since the nature and strength of people relations depend on many things representable by edges attributes. As a result, we move from a consideration of a usual unweighted graph as a mathematical model of social networks to the one with heterogeneous vertices and edges. Such networks are called *attributed,* and the paper is focused on their study.

There are many issues of the hour related to attributed network analysis such as ATNs' formation models; recovering those attributes that are missing; CD in attributed SNs that utilizes knowledge on a network nature contained in its attributes associated with edges and vertices; interpretation of CD-results, and many other related issues. The later problems were studied in [1, 2], where a notion of a multi-layer attributed network (ATN) was introduced, explored and applied in SN Analysis (SNA). This paper is dedicated to the investigation of the first problem related to ATNs' formation modelling. Its solution is of high importance since this allows a deeper understanding of the nature of social and other ATNs and, accordingly, modelling and solving problems of our interest, such as CD, clustering, restoring networks etc.

2.1. Social Networks' Peculiarities

Networks depicted people relationships are called social (SNs). In a graph associated with the network, users of the network are representable by vertices and the users' contacts - by its edges or arcs. In the graph, people heterogeneity is accumulated in attributes of the vertices, while edges' attributes depict a variety of their relationships.

An SN is defined as the following:

Definition 1 *[3] A social network is a hybrid graph represented in the form of:*

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

with a set of vertices V (social network's users), is a set of edges E (the users' relationships), Λ and Λ' accumulate data on attributes for each vertex $v \in V$ and edge $\{u, v\} \in E$, respectively.

Due to SNs are formed based on human relationships, they have specific features that are not characteristic of other networks, such as production, transport, technological, biological ones, and so on. The specifics are caused by the necessity of humans to communicate. Among them are sparsity, low diameter and average shortest path, while the average clustering coefficient is high. Also commonly, there are a few highly sociable people with plenty of contacts, representable in an SN by high degree vertices (*hubs*). In contrast, there exist numerous unsociable humans with rare connections depicted by vertices of low degree. It results in a vertex degree distribution (VDD) with a "heavy" tail. Another crucial feature of SNs is in the presence of relatively dense subgraphs (communities). Concepts "small-world networks" and "scale-free networks" accumulate the listed properties [4, 5, 6, 7].

Definition 2 *[6]. A small-world network is a graph* G*, where the distance* $dist(u,v)$ *between two randomly chosen vertices* $u, v \in V$ *grows proportionally to the logarithm of its order n network, i.e.*

$$dist(u,v) \propto log\ n. \tag{2}$$

Features of small-world networks are:

- **main** are high $\overline{CC}$ and small $\bar{l}$;

- **additional** are: a) the presence of many cliques and near-cliques caused by high $\overline{CC}$;

b) existence of a relatively short path between most of the pairs of vertices caused by small $\bar{l}$;

Definition 3 [7] *A network is called scale-free if its VDD is power-law, i.e., a probability* $p(m)$ *of appearing a vertex of degree* m *connections is approximated as the following:*

$$\exists \alpha \in R^1 : p(m) = P(d_u = m) \sim m^{-\gamma}, \forall u \in V. \tag{3}$$

Due to the listed features, SNs are of interest to researchers [5, 8, 9]. For instance, when SNs are modelled, the investigations primarily concentrate on reproducing small-world peculiarity (2) and scale-free graphs (3). Much fewer results are on modelling networks with clusters and communities, and even fewer papers consider an issue of simulating all the listed properties of SNs [5, 6, 7]. The real interest is riveted to the area of community detection (CD) [8, 10, 11, 12].

Communities may overlap or not. That is why we consider two types of vertex divisions.

Definition 4 *A partition of a network is its vertex division*

$$\mathcal{C} = \{C_k\}_{k \in [L]} : \tag{4}$$

$$\bigcup_{k=1}^{L} C_k = V, \tag{5}$$

where vertex clusters $C_k \subset V, k \in [L] = \{1, ..., L\}$ are pairwise disjoint.

Normally, CD assumes designing a network partition. Two classes of the partitions will be considered – a vertex partitions into communities being an output of any CDA and a network partitions into clusters associated with a certain value of the discrete vertex attribute. By combining all such partitions for all the vertex attributes, a resulting network cover is derived.

Definition 5 *A network's cover is a division (4) of the network* vertices *satisfying (5).*

Let [13] $E(C, C')$ be an edge set between vertex clusters $C, C' \subseteq V$. For $C \subseteq V$, let $G[C]$ be a subgraph G *induced* by C. For the network partition (4), the following notation will be used:

$$\mathcal{C}^* = \{C_l\}_{l \in [L^*]} \text{ is a partition } G \text{ into communities,} \tag{6}$$

$$|C_l| = |G[C_l]| = n_l, l \in [L^*], \sum_{l \in [L^*]} n_l = n. \tag{7}$$

For a network G, assume that we a given by attributes of vertices and (or) edges besides the standard information regarding its vertex set V and edge set E. Thus, the network is *decorated* [5,14] by these attributes and therefore belongs to the class of ATNs [15]. The question arises how this additional information on the attributes can be used to understand deeper the nature of the

network, in particular, is it helpful in CD? Our idea is that communities derived by CDAs are closely related to a single vertex attribute of an ATN or to their combination of these attributes.

3. Results and Discussion

3.1. ATN Formation Models

In the current section, we will consider the issue of forming ATNs. We will be interested in how an ATN (1) is created, provided attributes of vertices and their values Λ are known. Thus, we investigate ways of forming an edge set E, the attributes and matrix Λ'. Here, we assume that the network edges are formed only based on the similarity of the vertices' attributes. Three ways to solve the problem will be outlined resulted in three ANT models. All of them can be seen as models of SNs.

3.1.1. Model I - an association network model

As introduced in [14], *an association network* G^a is an ATN, where any identical value of an attributes results in creating an edge. Indeed, even if people do not know each other but have a common interest, they are related to each other to some extent. Maintaining the contact, in this case, does not cost anything since this contact is virtual.

With any activity/interest (AI), let us associate a network G^k, $k \in [K]$. Assume that networks G^k and $G^{k'}$ ($k \neq k'$) corresponding to different attributes are formed independently. In terms of [14], the weighted network $G^w = G^a$ can be represented as a weighted network sum (WNS)

$$G^w = \sum_{k \in [K]} w^k \cdot G^k. \tag{8}$$

of K subnetworks being a union of complete graphs:

$$G^k = \underset{l \in [L_k]}{\cup} K_{n_l^k}, k \in J_K. \tag{9}$$

Thus, G^a is a cover of K partitions of V by a disjoint union of complete graphs. Weights of the network are defined by (8), (10), where

$$w_{ij}^{ak} = \nu(k) \cdot W^{lk} \text{ if } \{v_i, v_j\} \in E, \text{ otherwise } 0 \ (i, j \in [n]), \tag{10}$$

$\nu(k)$ is a normalized factor of G^k making its weight equal to 1,

$$\exists l \in [L_k] : v_i, v_j \in AC_l^k. \tag{11}$$

3.1.2. Model II: an attributed Erdös-Rényi random graph-based model

Assume that, for the formation of an edge, the presence of identical attribute values is necessary but not sufficient. Likewise Model I, we form the network G^w by formula (8). Edges of the auxiliary G^k are formed with a probability p_l^k between two vertices v_i, v_j with a value at_l^{nk} of AT^{nk}. Thus, the network G^k will be a vertex partition by Erdös-Rényi Random Graphs (ERRGs) [14]:

$$G^k = \underset{l \in [L_k]}{\cup} ERRG(p_l^k, n_l^k), k \in [K]. \tag{12}$$

The accumulated network $G^w = G^{wII}$ will be a cover, where K partitions by ERRGs are overlapped. Weights of edges of a subnetwork G^k are defined as follows

$$w_{ij}^{k(II)} = \nu'(k) \cdot W^{II,lk} \text{ if } \{v_i, v_j\} \in E, \text{ otherwise } 0 \ (i, j \in [n]), \tag{13}$$

where

$$\exists l \in [L_k] : v_i, v_j \in AC_l^k, \{v_i, v_j\} \in E; \tag{14}$$

$$v'(k) = \left(2 \cdot \sum_{l \in [L_k]} W^{II,lk} \cdot m_l^k \right)^{-1}, k \in [K].$$

Model II simulates a real situation of establishing human contacts when a group of people are formed simultaneously. They have no opportunity to analyze who of the group members is the most interesting person. As a result, contacts arise randomly, and then they become permanent only if there are indeed commonalities of the people interests.

3.1.3. Model III - an attributed Barabasi-Albert preferential attachment-based model

In contrast to Model II, we simulate a situation when the team is formed gradually in the next model. Respectively, new people can analyze who is more influential and interesting in the group from his perspective and establish contacts.
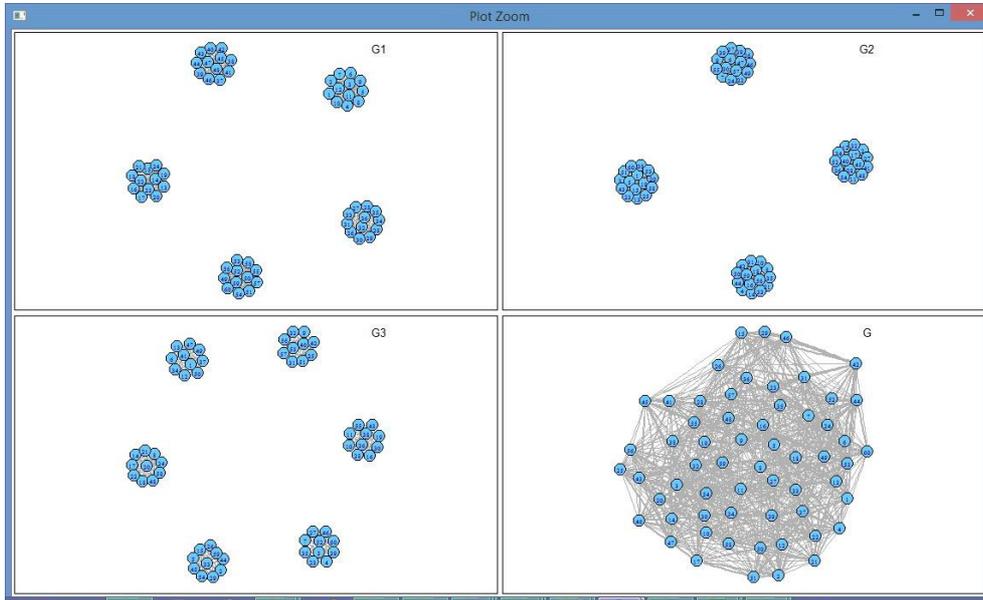

Figure 1: Model I: the layers $G^1 - G^3$ and the resulting weighted association network G^d made from them

$G^w = G^{wIII}$ will also be a WNS (8) of subnetworks G^k, $k \in [K]$. The networks $\{G^k\}_k$ are formed consecutively as k increases with respect to their weights. For $k \in [K]$, the edge set of the network G^k is formed within clusters of vertices with the identical value of AT^{nk} consecutively by $i \in J_n$ with probabilities that depend on degrees $\{d_{i'}^k\}_{i'}$ of all preceding vertices $v_{i'}$ $(i' < i)$ as well as on parameters p^k, p'^k $(p^k \leq p'^k)$ for new and earlier established contacts, correspondingly. The model uses Barabasi-Albert Preferential Attachment Model [7] and generalizes it onto ATNs. Another way of generalization is to form each network G^k as follows: the disjoint subsets of vertices to link according to the preferential attachment. The resulting subgraphs will create a vertex partition $\mathcal{AC}^k$. Then, these all are united into a multi-layer cover $\mathcal{AC}$. In contrast with the two above models, the network's layers will be dependent regardless $p^k = p'^k, k \in [K]$ (Model III.1) or the one where $\exists k \in J_K : p^k < p'^k$ (Model III.2).

Model III.1 induces a vertex partition by Barabasi-Albert random Graphs (BARGs). G^k can be represented likewise (9), (12) yielding:

$$G^k = \underset{l \in [L_k]}{\cup} BARG(n_l^k, \beta_l^k), \; k \in [K],$$

(15)

where β_l^k is the power of preferential attachment in AC_l^k, $l \in [L_k], k \in [K]$.
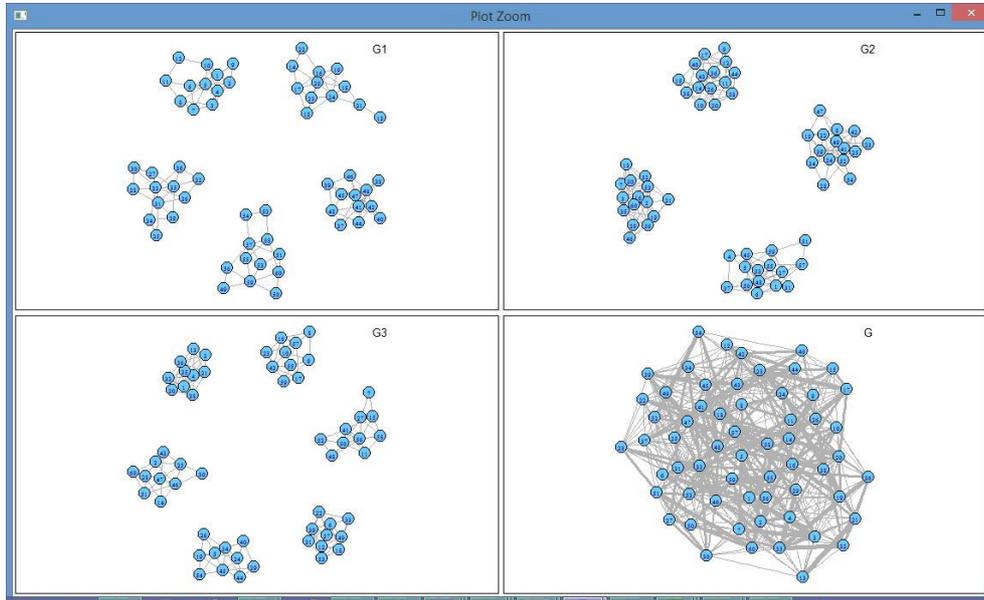

Figure 2: Model II: the layers $G^1 - G^3$ and the resulting weighted network G^{wII} made from them
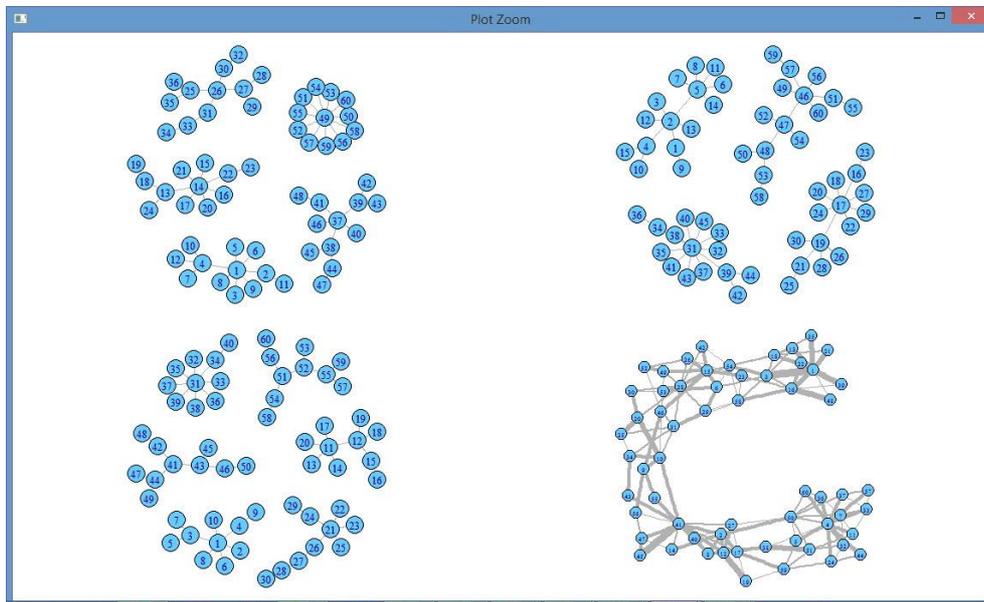

Figure 3: Model III: the layers $G^1 - G^3$ and the resulting weighted network G^{wIII} made from them

3.2. ATN Simulation

The ATN Models I-III introduced in Section 4 we illustrate by examples of simulations in the IGraph implemented in R. The first one is a simulation of an association network G^a (Model I), the second one demonstrates Model II (a network G^{wII}), and the last one shows Model III (a network G^{wIII}).

These networks have the following common characteristics: the order is $n = 60$, the number of the vertex attributes is $K = 3$, the vertices are randomly partitioned into $\{L_k\}_k = \{5, 4, 6\}$ attribute clusters of equal size $(n_l^k)_{l \in [L_k], k \in [K]} = ([n^k]^{L_k})_{k \in [K]} = (12^5, 15^4, 10^6)$.

3.2.1. Examples of Models I-III simulation

The weights of the ATNs are $\overline{W}^{II} = (0.5, 0.3, 0.2)$.

Example 1 - Model I simulation

G^a is the weighted network sum of vertex partitions by $5, 4, 6$ complete graphs, respectively (see (8), (9)): $G^a = 0.5 \cdot G^1 + 0.3 \cdot G^2 + 0.2 \cdot G^3$, e.g., $G^1 = \underset{l \in [5]}{\cup} K_{12}$. In Figure 1, one can see the graphs G^1, G^2 and G^3, which are partitioned by a disjoint union of complete graphs, along with the corresponding association network G^a.

Example 2 - Model II simulation

According to (8) and (12), G^{wII} is the following weighted network sum of vertex partitions by random graphs: $G^{wII} = 0.5 \cdot G^1 + 0.3 \cdot G^2 + 0.2 \cdot G^3$. The result of the simulation with parameters $(p_l^k)_{l \in [L_k], k \in [K]} = ((p^k)^{L_k})_{k \in [K]} = (0.3^5, 0.3^4, 0.5^6)$ is depicted in Figure 2. Here, for example, $G^1 = \underset{l \in [5]}{\cup} ERRG(p^l, n^l) = \underset{l \in [5]}{\cup} ERRG(0.3, 12)$ and G^{wII} is overlapping of three such partitions by random graphs.
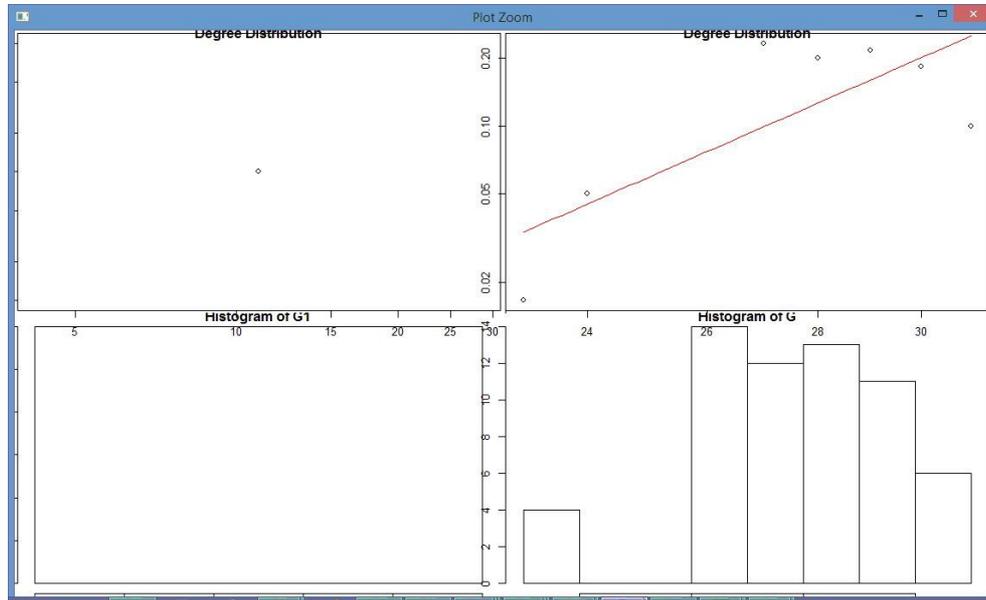

Figure 4: Model I. Top two - in log scale, a comparison of degree distribution with power-law in G^1 (left), G^a (right); Bottom two – the degree distribution in G^1 (left) and G^a (right).

Example 3 - Model III simulation

Similarly, for the network G^{wIII}, which is simulated according to Model II, we chose a linear preferential attachment model $((\beta_l^k)_{l \in [L_k], k \in [K]} = (\beta)_{l \in [L_k], k \in [K]}, |\beta| = 1)$. G^{wIII} is formed by (8), (15) and shown in Figure 2. Here, for instance, $G^1 = \underset{l \in [5]}{\cup} BARG(12, 1)$ is a disjoint union of BARGs. At

the same time, G^{wIII} is a connected graph being a cover by overlapping interconnected graphs formed from G^1, G^2, G^3 which are formed bases on vertex attributes of consecutively arising vertices.
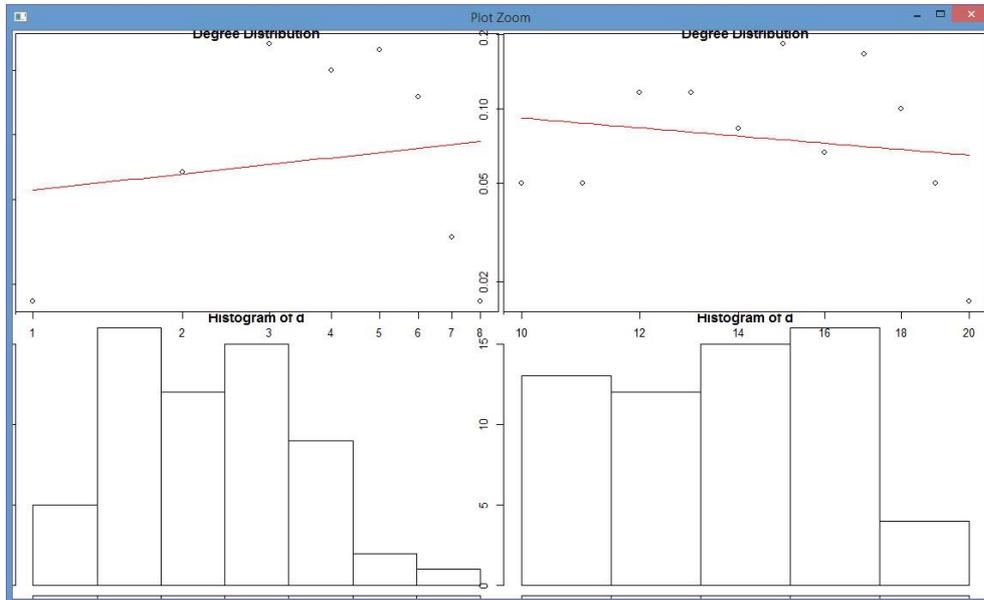

Figure 5: Model II. Top two - in log scale, a comparison of degree distribution with power-law in G^1 (left), G^{wII} (right); Bottom two – the degree distribution in G^1 (left) and G^{wII} (right).

3.2.2. Attributed networks Models Comparison

The purpose of this section is to identify to which extend the proposed network models reproduce social networks. For this, we analyze the listed above properties of SNs and make a comparison of them.
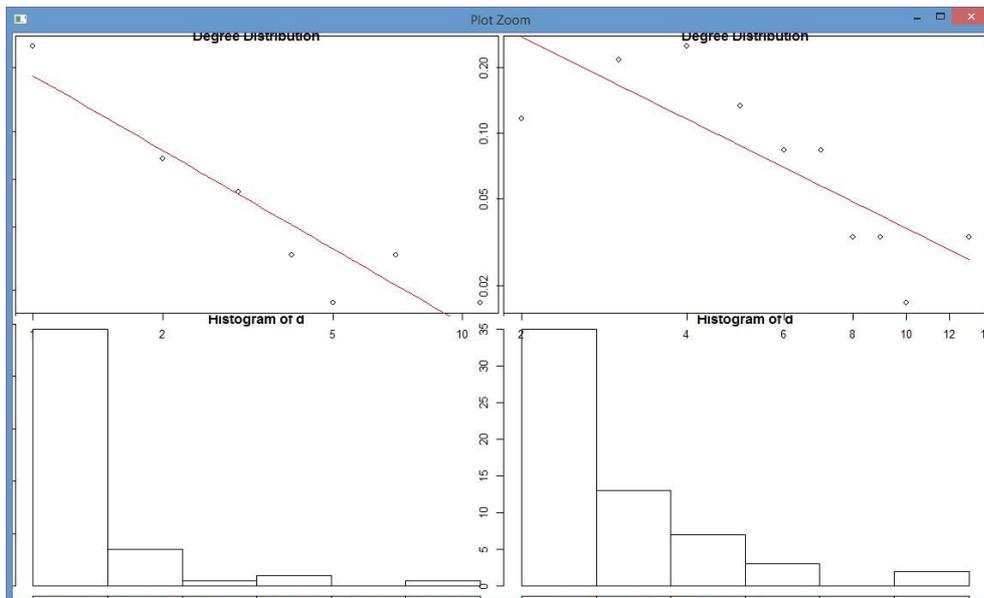

Figure 6: Model III. Top two - in log scale, a comparison of degree distribution with power-law in G^1 (left), G^{wIII} (right); Bottom two – the degree distribution in G^1 (left) and G^{wIII} (right).
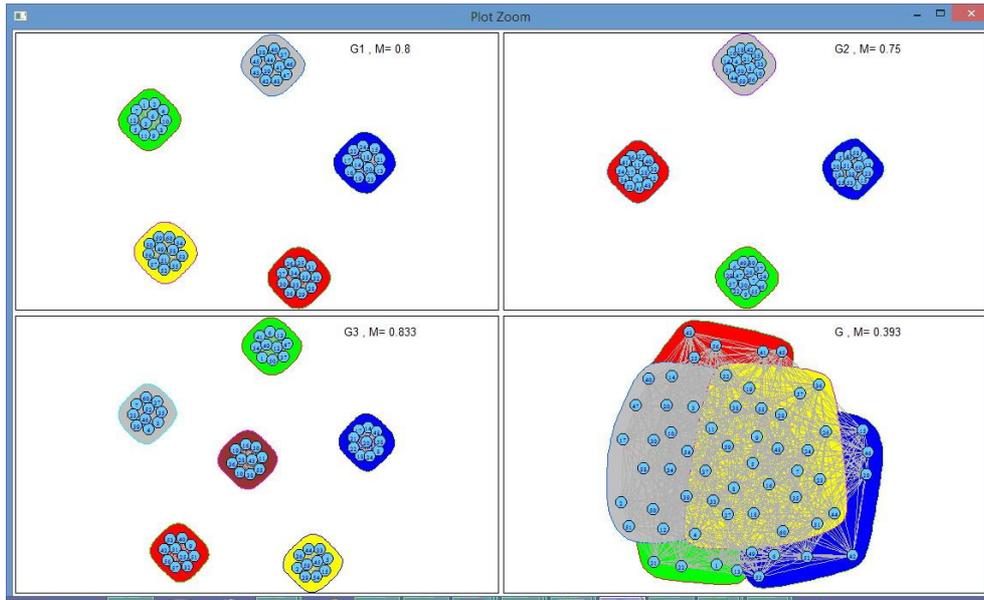
Figure 7: Model I: modularity values for G^a and its subnetworks $G^1 - G^3$

We conduct CD for all three simulated networks. Table 1-3 shows the obtained values of modularity along with other metrics related to SNs. One can see that modularity is high enough for subgraphs $G^1 - G^3$, namely, $M \in [0.747, 0.833]$. At the same time, for the resulting network, this value is much lower due to overlapping the above subgraphs, namely, $M \in (0.393, 0.640)$. Only for the third model, the modularity is still high enough and indicates a clear community structure. The CD results on G^a, G^{wII}, and G^{wIII} are illustrated in Figures 7-9.
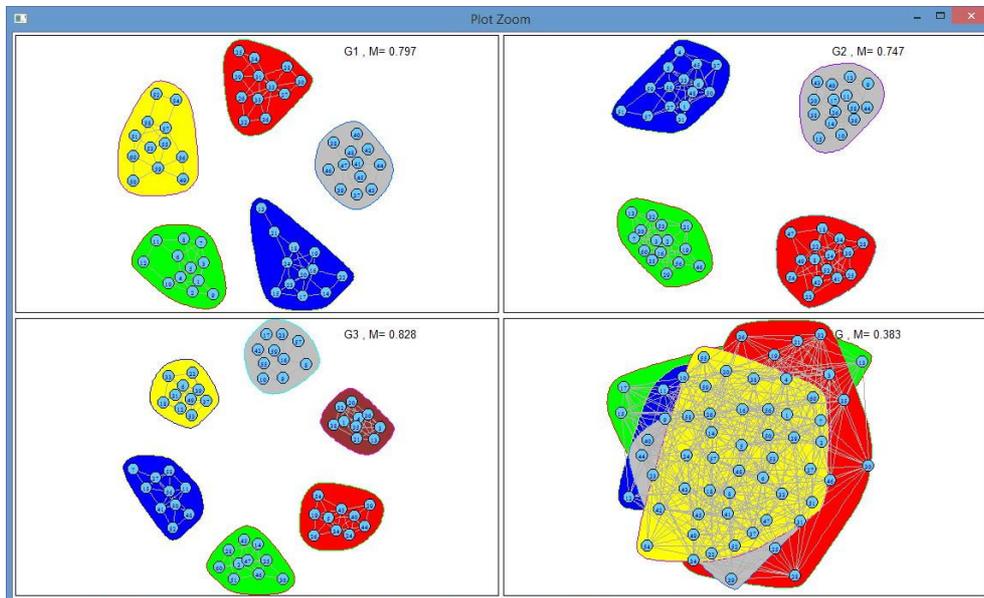

Figure 8: Model II: modularity values for G^{wII} and its subnetworks $G^1 - G^3$
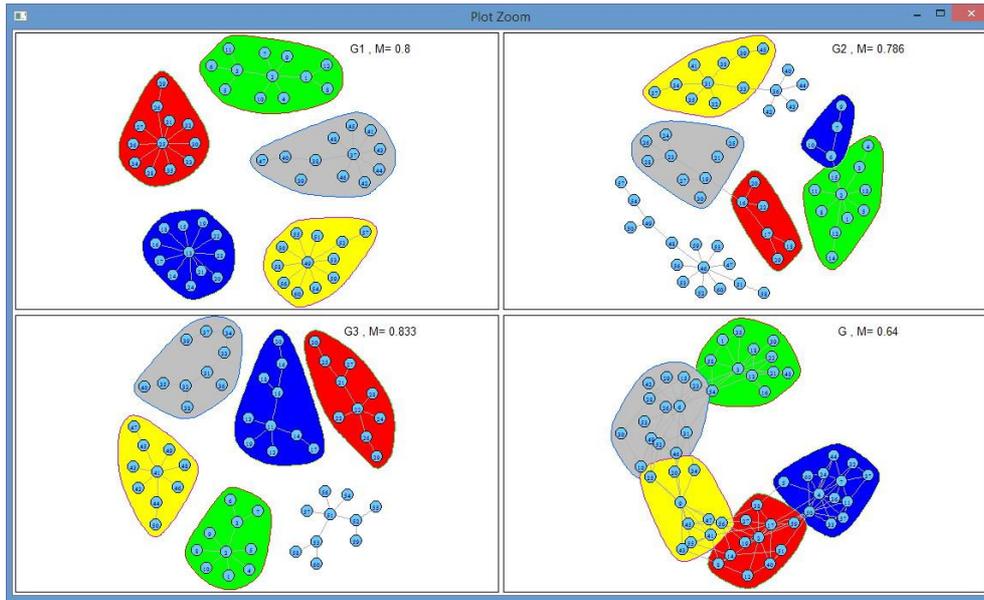
Figure 9: Model III: modularity values for G^{wIII} and its subnetworks $G^1 - G^3$

Table 1

Model I: the metrics in G^a

i	metric	G^1	G^2	G^3	G^{wIII}
1	n	60	60	60	60
2	m	330	420	270	851
3	δ	0,186	0,237	0,153	0,481
4	$\bar{l}$	1	1	1	1,519
5	$\overline{CC}$	1	1	1	0,567
6	M	0,8	0,75	0,833	0,393

Table 2

Model II: the metrics in G^{wII}

i	metric	G^1	G^2	G^3	G^{wIII}
1	n	60	60	60	60
2	m	128	214	154	446
3	δ	0,072	0,121	0,087	0,252
4	$\bar{l}$	1,7	1,507	1,441	1,783
5	$\overline{CC}$	0,346	0,522	0,534	0,34
6	M	0,798	0,747	0,832	0,403

Table 3

Model III: the metrics in G^{wIII}

i	metric	G^1	G^2	G^3	G^{wIII}
1	n	60	60	60	60
2	m	55	56	54	145
3	δ	0,031	0,032	0,031	0,082
4	$\bar{l}$	2,091	2,776	2,407	3,975
5	$\overline{CC}$				0,476
6	M	0,8	0,786	0,833	0,64

Table 4
Comparison of Models I-III

metric	G^a	G^{wII}	G^{wIII}
δ	0	0,5	1
$\bar{l}$	1	1	0
$\overline{CC}$	1	0	0,5
M	0,5	0,5	1
power-law distribution	0	0	1
Total	2,5	2	3,5

Here, density $\delta = \dfrac{2m}{n(n-1)} \ll 0.5$, where $n = |V|$, $m = |E|$; an average shortest path length

(ASPL) $\bar{l} = \dfrac{1}{m} \sum_{\{u,v\}\in E} l(u,v)$, where $l(u,v)$ is the shortest path length between $u,v \in V$; the average

clustering coefficient (ACC) $\overline{CC} = \dfrac{1}{n}\sum_i CC_i$, where $CC_i = \dfrac{2d_i'}{d_i(d_i-1)}$ is the local clustering

coefficient of $v_i \in V$, d_i, d_i' is the v_i-degree and number of edges within a neighbourhood of v_i;

$M = 0.5\sum_l \sum_{u,v\in C_l} \left(a_{u,v}^w - d_u d_v / 2\right)$ is the modularity, where d_u is the degree of $u \in V$.

The densities of the networks G^a, G^{wII}, G^{wIII} are a bit lower than the one of a sum of $G^1 - G^3$ one. The network is the densest with $\delta(G^a) = 0.481$, while the network G^{wIII} is the sparsest having $\delta(G^{wIII}) = 0.082$. Besides, G^{wII} is rather sparse and demonstrates $\delta(G^{wII}) = 0.252$. In all these networks, the ASPL-metric is low, and its values fluctuate around 1.6. The only exception is the network G^{wIII} utilizing preferential attachments, namely, $\bar{l}(G^{wIII}) = 3.975$. On the other hand, the ACC-value in this network is quite high, namely, $\overline{CC}(G^{wIII}) = 0.476$. At the same time, in the second network G^{wII}, the value is less - $\overline{CC}(G^{wII}) = 0.34$.

In the last part of the computational experiment, we analyzed the power-low distribution of vertex degrees in our modelled networks. Figure 4-6 shows that in the resulting networks G^a and G^{wII}, the degree distribution differs significantly from the distribution of the subnetworks, particularly in G^1, and from power-law. Only the latest model G^{wIII} shows in Figure 6 behaviour similar to a power-law distribution. In Figure 6 from the two later plots, it is clearly seen the havier right tail in the accumulated network G^{wIII} than the one in G^1. The two early plots demonstrate fitting the degree distribution by a linear function after switching to a logarithmic scale. The multiple determination coefficient of the linear regressions gets values $R^2(G^1) = 0.646$ and $R^2(G^{wIII}) = 0.761$, respectively. This says that an exponential function fits good the achieved degree distribution of the presented network models. The exponential approximation curve better fits the preferential attachment-based Model III. We can conclude that the AN G^{wIII} is the closest to a SN. It has the highest overall rank for the evaluated SN metrics (see Table 4). In the table, the rank 0 corresponds to the worst-case and 1 - to the best. Also, the accumulation of three partitions by BARGs results in an AN closer to an SN than the graph partitions itself. Also, note that in terms of human communication, the networks G^{wII} and G^{wIII} simulate contrasting situations, where contacts are established absolutely by accident or made fully intend. The real situation is a combination of these two. That is why we propose one more network model $G^w = W^2 G^{wII} + W^3 G^{wIII}$, where

$W^2 + W^3 = 1$, $W^2, W^3 \geq 0$. We expect that G^w is can take advantages of its component, such as the ASPL decreases thanks to G^{wII}, while the ACC rises due to the presence of G^{wIII}.

4. Conclusions

In the paper's scope, decomposition of social and other attributed networks (ATNs) into subnetworks related to single vertex attributes is studied as a way of their modelling. A number of models of ATNs with a finite number of discrete-valued vertex attributes is presented. They utilize famous graph models such as complete, Erdös-Rényi and Barabasi-Albert random graphs. Experimentally, it is shown that the proposed ATN models, particularly the Barabasi-Albert graph-based model, are enable to reproduce most social networks peculiarities. It is progress in comparison with known so far networks models.

5. References

[1] B. Farzad, O. Pichugina, L. Koliechkina, Multi - Layer Community Detection, in: 2018 International Conference on Control, Artificial Intelligence, Robotics Optimization (ICCAIRO), 2018, pp. 133–140. doi:10.1109/ICCAIRO.2018.00030.

[2] Pichugina, B. Farzad, A Human Communication Network Model, in: ICTERI, 2016.

[3] Z. Lin, X. Zheng, N. Xin, D. Chen, CK-LPA: Efficient community detection algorithm based on label propagation with community kernel, Physica A: Statistical Mechanics and its Applications 416 (2014) 386–399. doi:10.1016/j.physa.2014.09.023.

[4] M. E. J. Newman, The Structure and Function of Complex Networks, SIAM Review 45 (2003) 167–256. doi:10.1137/S003614450342480.

[5] E. D. Kolaczyk, Statistical Analysis of Network Data, Springer Series in Statistics, Springer New York, New York, NY, 2009. doi:10.1007/978-0-387-88146-1.

[6] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442. doi:10.1038/30918.

[7] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (2002) 47–97. URL: http://link.aps.org/doi/10.1103/RevModPhys.74.47. doi:10.1103/RevModPhys.74.47.

[8] M. Girvan, M. E. J. Newman, Community Structure in Social and Biological Networks, in: Proceedings of the National Academy of Sciences of the United States of America, volume 99/12, 2002, pp. 7821–7826.

[9] E. Cuvelier, M.-A. Aufaure, Graph Mining and Communities Detection, in: Business Intelligence, number 96 in Lecture Notes in Business Information Processing, Springer Berlin Heidelberg, 2012, pp. 117–138.

[10] S. Fortunato, Community detection in graphs, Physics Reports 486 (2010) 75–174.

[11] A. Lancichinetti, S. Fortunato, Community detection algorithms: a comparative analysis, Physical Review. E, Statistical, Nonlinear, And Soft Matter Physics 80 (2009) 056117:01–056117:17.

[12] M. E. J. Newman, Communities, modules and large-scale structure in networks, Nature Physics 8 (2012) 25–31. doi:10.1038/nphys2162.

[13] C. Staudt, Y. Marrakchi, H. Meyerhenke, Detecting communities around seed nodes in complex networks, in: 2014 IEEE International Conference on Big Data (Big Data), 2014, pp. 62–69. doi:10.1109/BigData.2014.7004373.

[14] J. Bruning, V. A. Geiler, I. S. Lobanov, Spectral Properties of Schrodinger Operators on Decorated Graphs, Mathematical Notes 77 (2005) 858–861. doi:10.1007/s11006-005-0086-z.

[15] Y. Zhou, H. Cheng, J. Yu, Clustering Large Attributed Graphs: An Efficient Incremental Approach, in: 2010 IEEE 10th International Conference on Data Mining (ICDM), 2010, pp. 689–698. doi:10.1109/ICDM.2010.41.

[16] P. Erdös, A. Rényi, On random graphs, I, Publicationes Mathematicae (Debrecen) 6 (1959) 290–297. URL: http://www.renyi.hu/~p_erdos/Erdos.html#1959-11.