Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 3 Issue 4, April - 2016

Comparative Analysis of the Structural and Weighted Properties in Albanian Social Networks

Eva Noka (Jani)
Department of Applied Mathematics
Faculty of Natural Sciences, University of Tirana
Albania
eva.jani@fshn.edu.al

Fatmir Hoxha
Department of Applied Mathematics
Faculty of Natural Sciences, University of Tirana
Albania
fatmir.hoxha@fshn.edu.al

***Abstract*—In this paper, we perform a comparative analysis of two Albanian social networks, which are constructed using the data of scientific collaborations and the records of communications (calls and text messages) between mobile phone users. Both networks display similar structural properties in accordance with other social networks, such as a broad degree distribution, high clustering and a community structure. The link weights vary in a wide range indicating a strong heterogeneity of the tie strengths in both networks. Exploring the correlations between tie strengths and topology, we observe significant differences. The communities in the collaboration network are mostly associated with the weak ties, whereas strong ties act as bridges connecting different communities. On the contrary, in the communication network, the strong ties mainly reside within the communities and the weak ties play the important role for overall network connectivity. These findings suggest that despite the structural similarities, the networks under consideration are driven by different mechanisms of network tie formation determining qualitatively different roles of the weak and strong ties at both local and global levels.**

Keywords—social networks; tie strength; weighted networks; percolation

I. INTRODUCTION

A social network is a map of the social structure, where every node represents a person, group or organization and pairs of nodes are linked through some kind of relation. A relation may refer to acquaintance, kinship, friendship, scientific co-authorship or other. This representation provides not only the observation of individuals and their attributes, but the relationships between them, their global structure and dynamics. The main outcome of the extensive research to understand, measure and model the structure of human-driven real world networks, has been to reveal that real networks significantly differ from random networks and despite the inherent differences, they share many similar topological properties [1, 2, 3].

Nevertheless, all these properties assume the links to be equivalent to each other, whereas, in the real life, the relationships that people maintain have not the same intensity, importance and role [4, 5]. For this reason, the social networks are better described in terms of weighted networks, where each link ij carries a numerical value w_{ij} as a weight, which measures the strength of the tie [6, 7]. Exploring the tie strengths and their correlations with the topological structure helps in a better description of the hierarchies, organizational principles of the networks and the dynamics of many phenomena, such as communities' formation, information spreading and social influence [8, 9, 10, 11].

We explore in a comparative way to social networks based on the data of scientific collaborations of the Albanian researchers and the communication records of the mobile phone users from an Albanian cellular operator. We find that both networks display many characteristic features of other social networks, such as broad degree/weight distributions, high clustering coefficient and community structures.

In the networks with a community structure the ties can be distinguished by their position. A tie can either resides within a community or acts as a bridge connecting different communities. The intercommunity ties are characterized by a high betweenness centrality, defined as the number of shortest paths in the network passing through a given link [12]. This implies that intercommunity ties have a high control on the information flow in the network. On the other hand, depending on their weights, the ties may be strong or weak. Hence, exploring the link position-weight relationships in weighted networks provides an additional source to unveil the driving mechanisms of tie formation and reinforcement. If there should not any correlation between the topology and the link weight, the strength of a particular tie should be only related to the nature of the relationship between the two individuals and thus is independent of the network structure around that tie. This is known as the dyadic hypothesis. Many real networks are self-organized according to the global efficiency principle, meaning that the tie strengths are optimized to maximize the overall flow in the network. In this case, the strongest ties in the network should be mainly intercommunity ties having a high betweenness, whereas the ties inside the communities should be weak. Another hypothesis is known as “the strength of weak ties” and comes from sociology [4]. It predicts that the strength of a tie between two individuals increases with the overlap of their friendship circles, implying that the ties within communities tend to be stronger than the ones between them. This hypothesis may be seen as the opposite of the global efficiency principle.

The exploring of the weight-topology correlations in our networks suggest that the collaboration network is organized according to the global efficiency principle, whereas the weak tie hypothesis is verified in the mobile phone communication network.

In the next Section we describe the data sets and the construction of the networks. Then, we analyze some basic characteristic and find that both networks display a community structure. We further address correlations between tie strengths and the structure of the networks. Exploring the link percolation properties we show that the strong ties have a global important role for the connectivity of the collaboration network, while this role in the communication network is played by the weak ties.

II. Data sets

A. *Scientific collaborations*

The first dataset contains all papers published in the Bulletin of the Faculty of Natural Sciences of the University of Tirana, from the No. 1 in the year 2004 to the No. 18 in 2015. We construct a collaboration network (ACN), which is a simple graph, where authors are identified with nodes and there is a link between a pair of nodes if the corresponding scientists have co-authored at least one paper. The network has $N = 591$ nodes, $L = 1\,115$ links. The average degree is $\langle k \rangle = 3.77$, which means that each scientist collaborates on average with $3-4$ other scientists. The largest connected component (LCC) amounts to 66.2% of the total number of nodes.

To measure the intensity w_{ij} of the collaboration between two scientists i and j, we use the definition of the link weight introduced by Newman in [7] as: $w_{ij} = \sum_p \frac{1}{n_p - 1}$, $i \neq j$, p is the set of papers where scientists i and j have collaborated and n_p is the number of co-authors of paper p. The motivation of this definition is based on a reasonable assumption that the collaboration intensity of two scientists on a paper with many other authors is less than on a paper with fewer authors. Each scientist who collaborates on a paper p with n_p co-authors divides his/her time evenly (on average) among $n_p - 1$ other authors, hence the intensity of his/her collaboration with each of them is $1\,/\,(n_p - 1)$.

B. *Mobile phone communications*

The second dataset contains all the records of calls and text messages (SMS) exchanged between the mobile phone users (anonymized) of a single Albanian operator during a period of one month. To give a representation of the social ties close to the real one, we construct a mutual communication network (MCN), where two users i and j are connected with an undirected link if there had been at least one reciprocated pair of communications (call or SMS) between them, i.e. i has initiated a call or SMS to j and vice versa [9]. The network has $N = 39\,535$ nodes and $L = 47\,430$ links. The average degree is $\langle k \rangle = 2.4$, meaning that a user communicates reciprocally with about 2-3 persons averagely. The LCC constitutes 69.7% of the total number of nodes. We quantify the weight of a link e_{ij} by the total number of communication events (calls or SMS) occurred between the users *i* and *j* over the studied period.

III. Results

A. *Basic characteristics*

The basic characteristic of the nodes in a network is the degree distribution denoted as $P(k)$, which gives the probability that a randomly chosen node has degree k. Panels A and B in Fig. 1 show the cumulative degree distributions $P_>(k)$ of the networks ACN and MCN, respectively. The skewed degree distributions with fat tails in both networks reflect the heterogeneous structure of the networks, a common feature for complex networks [13]. For the collaboration network ACN, this means that while most scientists have only a few collaborators, a small minority collaborates with many others. One scientist collaborates maximally with 42 others or more than 7% of the total number of authors. Although the communication network MCN has a large size, its maximum degree is $k_{max} = 77$, indicating that the hubs are few. The reciprocal communication criterion for the construction of MCN filter out the abnormal large nodes, which in most cases are associated with business-like subscriptions, customer service lines, etc., hence the resulting mutual network is dominated by trusted interactions.

Link weights display a broad distribution in both networks (Fig. 2 A, B), reflecting the large variability of the interactions’ intensity between individuals in their professional relationships as well as in their friendship relations. We observe that most ties are weak, whereas a tiny fraction of them are strong.

The local clustering coefficient C_i measures the density of links in the immediate neighborhood of the node i [14]. It is defined as:

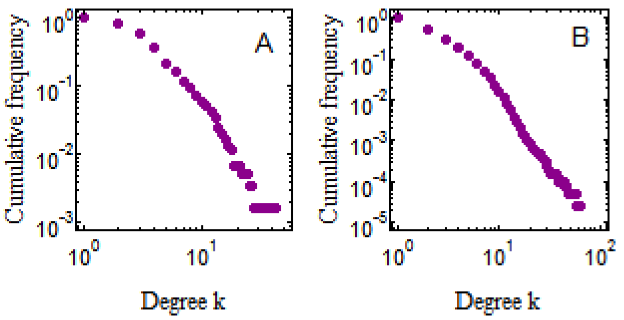

Fig. 1. *Cumulative degree distributions for the collaboration network ACN (A) and for the communication network MCN (B).*

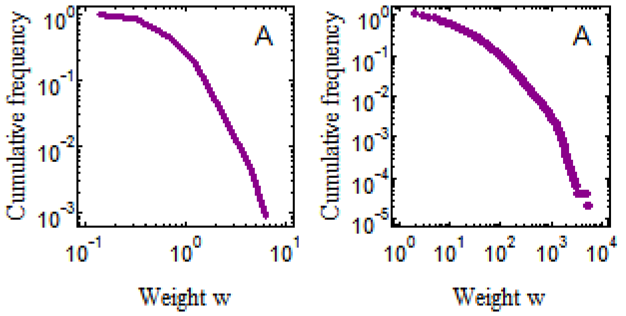

Fig. 2. C*umulative weight distributions for the collaboration network ACN (A) and for the communication network MCN (B).*

$$C_i = \frac{2t_i}{k_i\,(k_i-1)} \qquad (1)$$

where $t_i$is the number of triangles connected to node i. In general, in social networks, the clustering coefficient is considerably higher than for a random network with similar number of nodes and links [3]. We have plotted (Fig. 2 A, B) the average local clustering coefficient $\langle C|k\rangle$ of all the nodes with the same degree k as a function of k, for both networks ACN and MCN. For each network, we compare the results with those taken from a configuration model. This reference model is obtained from the original network by preserving its degree sequence, but the links are rewired pairwise randomly to remove the local structural correlations present in the network. We observe different behaviors of $\langle C|k\rangle$ in the original networks and in their randomized counterparts. The small degree nodes have a much higher $\langle C|k\rangle$ in the empirical networks than in the link-randomized networks, meaning that the local structure around the small degree nodes is denser than expected by chance. While in the link-randomized networks, $\langle C|k\rangle$ is independent of k, in the original networks $\langle C|k\rangle$ shows a decreasing trend with the increase of k. This means that the empirical networks display a hierarchical structure [3], where the small degree nodes are part of small dense communities, while large degree nodes link different communities to each other. These observations suggest that the links in our empirical networks are not random. On the contrary,

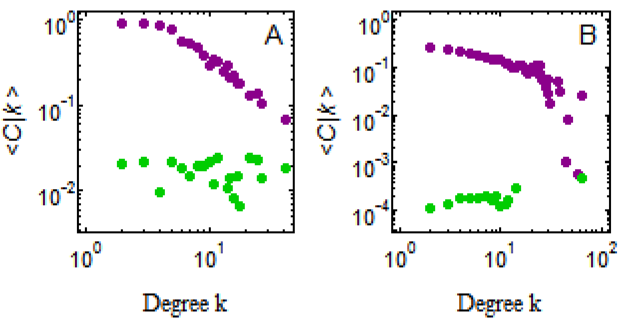

Fig. 3. *Clustring coefficient. (A) Average clustering coefficient for the ACN (purple dots) and the randomized network (green dots). (B) Average clustering coefficient for the MCN (purple dots) and the randomized network (green dots).*

the nodes arrange themselves locally in well-organized structures, reflecting the tendency of the individuals to be organized in tightly knit groups according to their interests, the needs for social cohesion and access to resources and new information [4, 14, 15] under time and cognitive constraints [16, 17, 18]

B. Weight-topology correlations

The *intensity* of a subgraph is an important concept designed for studying the coupling between the network structure and tie strengths [onnela]. The intensity of a subgraph g with nodes v_g and links l_g is given by the geometric mean of its weights as:

$$I(g) = \left(\prod_{ij\in l_g} w_{ij}\right)^{1/|l_g|} \qquad (2)$$

We analyze the intensity distributions of the k-cliques (fully connected subgraph with k nodes) for $k = 3, 4$ and 5 in the largest connected components (LCC) of the empirical networks. For each of them, we set up a *weight permuted reference* [9, 19] generated by randomly reshuffling the weights in the network. This reference model removes weight correlations in the original network keeping the topology unaltered. The k-clique intensity distributions for the LCC of the empirical networks and their respective reference models are displayed in Fig. 3.

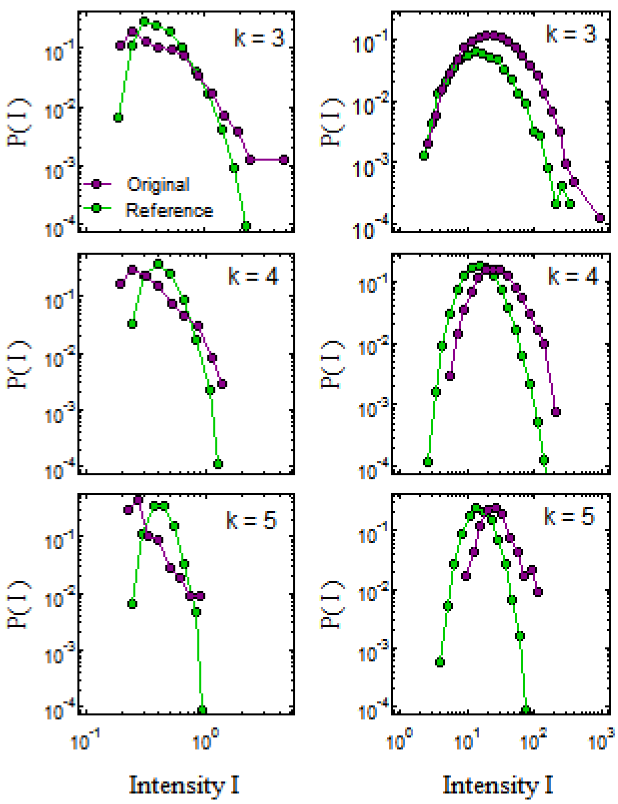

Fig. 4. *Clique intensity distributions. The probability distribution of cliques of oreder $k = 3,\ 4,\ 5$ in the LCC of the ACN (left column), MCN (right column) and their reference models. The reference distribution is an average over 100 realizations.*

In the collaboration network, we observe that for each order, most cliques have low intensity, while only a small number of cliques have high intensity. On the contrary, in the communication network, the intensity of cliques is considerably higher than in the randomized counterpart. The differences become more evident for larger cliques. Furthermore, the median k-clique intensities ($k = 3, 4, 5$) in the collaboration network are smaller than in the reference model, while the opposite is true in the communication network (Table I). The abundance of low-intensity cliques in the collaboration network indicates that dense network neighborhoods are mainly associated with the weak ties, whereas in the communication network, the intracommunity ties are seen to be stronger.

TABLE I. MEDIAN INTENSITY

Order k	Median intensity			
	Collaboration		***Communication***	
	O	***R***	***O***	***R***
3	0.333	0.395	22.915	14.675
4	0.287	0.404	26.066	14.972
5	0.267	0.410	24.858	14.981

Median intensity of cliques of oreder k, in the LCC of the original collaboration and communication networks (O) and their reference models (R).

A local topological property of a link is the *neighborhood overlap* [6]. For a link e_{ij}, the overlap O_{ij} represents the proportion of the neighbors common to its endpoints i and j. It is defined as:

$$O_{ij} = \frac{n_{ij}}{(k_i-1)+(k_j-1)-n_{ij}}\,, \qquad (2)$$

where k_i (k_j) is the degree of the node i (j) and n_{ij} is the number of the neighbors common to both nodes i and j. If two connected nodes i, j have no common neighbor, then $O_{ij} = 0$ and the link e_{ij} is a potential bridge between disparate communities. If all the neighbors of the nodes i and j are common to both, $O_{ij} = 1$, which means that the link e_{ij} is part of a single community.

We explore the average overlap $\langle O|w\rangle$ as a function of link weight w in the largest connected component of the ACN and MCN as shown in the panels A, B of Fig. 4 (purple dot-lines). The green dot-lines represent the average overlap $\langle O|w\rangle$ for the reference model. According to the dyadic hypothesis, the overlap is independent of the tie strengths in each of the link-weight randomized network. The different behavior of $\langle O|w\rangle$ in the original networks indicates that the dyadic hypothesis is not suited for our empirical networks, implying the existence of the correlations between the tie strengths and the network structures.

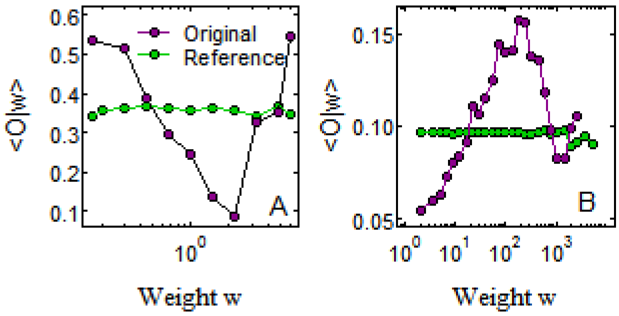

Fig. 5. *Average Overlap as a function of weight in the LCC of ACN (left) and MCN (right) and in their reference models.*

In the collaboration network, we observe that $\langle O|w\rangle$ decreases with the increase of the weight up to $w \approx 2.15$, which is followed by an increasing trend for larger values of w, constituting only 3% of the total links. This means that the ties within communities (with high overlap) are mainly weak, corroborating with the observations on the clique intensities. The intercommunity ties, which have a high flow control, are stronger, suggesting that the tie strength in the collaboration network is driven by the global efficiency principle. In contrast to the collaboration network, $\langle O|w\rangle$ increases with the increase of link weight for most links. The links with weight in the decreasing region of $\langle O|w\rangle$ constitute only 4.3% of the total number of links. In line with the weak tie hypothesis, the tie strength in our communication network is depended on the immediate neighborhood of the individuals involved, meaning that the more friends in common they have, the stronger becomes their tie.

C. Link percolation properties

The observed differences in the weight-topology correlations between the networks in consideration may have important implications on the roles that the weak and strong ties play in the connectivity of the networks in both local and global level. To understand this role we examine the link percolation properties in both the LCC of empirical networks. As in [9], in each network, we remove the links in decreasing and increasing order of weight (overlap), and calculate the fraction of nodes in the LCC of the remaining network R_{LCC}, as a function of the control parameter f, the ratio of removed links. The results are displayed in Fig. 5. The responses of the networks to the removal of links according to their overlap are similar, presenting a further evidence of their similar topological properties. Both networks shrink faster when the links are removed in increasing order of the overlap than by removing first the high overlap link s. This means that the low overlap links positioned between the communities are more important for the global connectivity of both networks. On the other hand, the networks display different behaviors to the removal of the links by their weights. The collaboration network disintegrates faster by removing the links in decreasing order of the weights, indicating that the strong ties are the most important links for the global

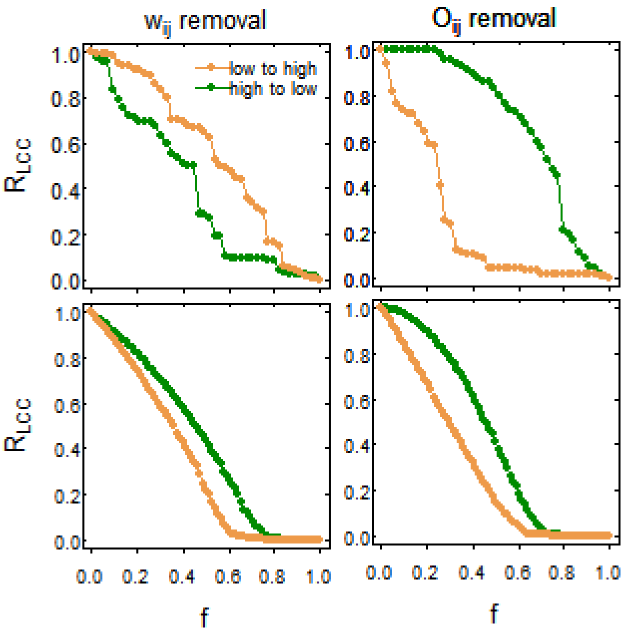

Fig. 6. *Link percolation in the LCC of the ACN (first row) and MCN (second row).*

connectivity, while the weak ties residing within communities are important for maintaining connected the community. In the communication network we observe that the removal of the weak ties first shrinks the network faster, indicating that the communities are linked with weak ties. At the local level, the strong ties play the role for the integrity of the communities.

D. Conclusions

Beside the topological characteristics of social networks, the tie strengths between individuals provide an additional source to a better understanding of their organizational principles and the driving mechanisms of tie formation and reinforcement. From this point of view, we analyzed two weighted social networks with data collected from the collaborations of Albanian scientists and mobile phone communications between the users of an Albanian operator. Both networks share many topological features observed in other social networks. They are characterized from the heterogeneity of both the degrees and tie strengths, high clustering coefficient and a community structure. Exploring the weight-topology correlations and the link percolation properties, we found that the networks have stark differences. In the collaboration network, we found that the tie strength depends on the global topology of the network, implying that the global efficiency principle is suited for this network. The ties between the scientists of different groups tend to be stronger and play important role for the overall connectivity of the network. This reflects the need of the scientists for avoiding the scientific isolation through interdisciplinary collaborations and the transmission of information in the network. On the contrary, in the communication network, the tie strength depends on the local structure around it in line with the weak tie hypothesis. The intercommunity ties are weak and act as bridges between communities and have the strength of keeping the network connected.

References

[1] S.H. Strogatz, “Exploring complex networks”, Nature, Vol. 410, 2001, pp. 268-276.

[2] M.E.J. Newman, “Networks. An Introduction”, Oxford University Press, New York, 2011.

[3] A.L. Barabási, “Network Science”, book available at http://barabasi.com/book/network-science.

[4] M. Granovetter, “The strength of weak ties”, American Journal of Sociology, Vol. 78, 1973, pp. 1360-1380.

[5] B. Wellman and S. Wortely, “Different strokes from different folks: community ties and social support”, American Journal of Sociology, Vol. 96, 1990, pp. 558-588.

[6] J.P. Onnela et al., “Structure and tie strengths in mobile communication networks”, Proc. Natl. Acad. Sci. USA, 2007, Vol. 104, No. 18, pp.7332-7336.

[7] M.E.J. Newman, “Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Phys. Rev., 2001, E 64, 016132.

[8] A.L. Hill, D.G. Rand, M.A. Nowak and N.A. Christakis, “Infectious disease modeling of social contagion in networks”, Plos Computational Biology, 2010, Vol. 6(11): e1000968.

[9] J.P. Onnela et al., “Analysis of a large-scale weighted network of one-to-one human communication”, New Journal of Physics 9 (2007), 179.

[10] D.J. Watts, “The “new” science of networks”, Annual Review of Sociology, 2004, .

[11] P. A. Grabowicz, J.J. Ramasco,E. Moro, J. M. Pujol, V.M. Eguiluz, ”Social features of online networks: The strength of intermediary ties in online social media. PLoS ONE, 2012, 7(1): e29358.

[12] P. Holme, “Edge overload breakdown in evolving networks”, Phys. Rev.,2002, E 66, 036119.

[13] A.L. Barabási, “Scale-free networks: A decade and beyond”, Science, 2009, Vol. 325, pp. 412-413.

[14] D.J. Watts and S.H. Strogatz, “Collective dynamics of ’small-world’ networks”, Nature, 1998, Vol.393, pp. 440–442.

[15] A.L. Barabási and R. Albert, “Emergence of scaling in random networks”, Science, 1999, vol. 286, pp. 509–512.

[16] R.I.M. Dunbar, “The social brain hypothesis”, Evolutionary Anthropology, 1998, 6(5), pp. 178–190.

[17] R. Hill and R.I.M. Dunbar, “Social network size in humans”, Human Nature-An Interdisciplinary Biosocial Perspective, 2003, Vol. 14, No. 1, pp. 53–72.

[18] S.G.B. Roberts, R.I.M. Dunbar, V.P. Thomas, T. Kuppens, “Exploring variation in active network size: Constraints and ego characteristics”, Social Networks 31 (2009) 138–146.

[19] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks”, Proc. Natl. Acad. Sci. USA, 2004, Vol. 101, No.11, pp. 3747-3752.