epl A Letters Journal Exploring the Frontiers of Physics

October 2008

EPL, **84** (2008) 28002
doi: 10.1209/0295-5075/84/28002

www.epljournal.org

Local assortativeness in scale-free networks

M. Piraveenan[1,2], M. Prokopenko[1(a)] and A. Y. Zomaya[2]

[1] *CSIRO Information and Communications Technology Centre - Locked Bag 17, North Ryde, NSW 1670, Australia*
[2] *School of Information Technologies, The University of Sydney - NSW 2006, Australia*

received 4 July 2008; accepted in final form 3 September 2008
published online 26 September 2008

PACS 89.75.Hc – Networks and genealogical trees
PACS 89.75.Da – Systems obeying scaling laws
PACS 89.75.Fb – Structures and organization in complex systems

Abstract – The level of assortative mixing of nodes in real-world networks gives important insights about the networks design and functionality, and has been analyzed in detail. However, this network-level measure conveys insufficient information about the local-level structure and motifs present in networks. We introduce a measure of *local assortativeness* that quantifies the level of assortative mixing for individual nodes in the context of the overall network. We show that such a measure, together with the resultant local assortativeness distributions for the network, is useful in analyzing network's robustness against targeted attacks. We also study local assortativeness in real-world networks, identifying different phases of network growth, showing that biological and social networks display markedly different local assortativeness distributions to technological networks, and discussing the implications to network design.

Copyright © EPLA, 2008

Introduction. – Assortativeness is the tendency observed in networks where nodes mostly connect with similar nodes. Typically, this similarity is interpreted in terms of degrees of nodes [1,2]. Many scale-free networks in real world show the tendency where highly connected nodes link with other highly connected nodes (that is, nodes mix assortatively). The reverse is also true in some networks, where highly connected nodes are more likely to make links with isolated, less connected nodes, *i.e.* to mix disassortatively. In both cases, the probability of creating a link depends on the degrees of both nodes. Averaging across the network, assortativeness quantifies the tendency for preferential association within the network [2,3].

Naturally occurring networks display various levels of assortative mixing, and a number of measures have been proposed to measure the level of assortativeness in such networks [1,2,4]. One such measure, and the one we follow in this letter, defines assortativeness as a correlation function in terms of degrees at the network level [2,5]. This, however, means that we can measure how assortative the network is as a whole, but not how locally-assortative are the individual nodes, or how do they contribute to the overall network assortativeness. To understand the local structure of the network, and the recurring local motifs in the network, it is, nevertheless, useful to analyze how the network assortativeness emerges from the individual assortative or disassortative tendencies of each node.

In this letter, we propose a measure of *local assortativeness*, which is a property of a single node and indicates how similar a node is to its neighbours. This similarity is interpreted in node degrees. We define local assortativeness as an individual node's contribution to the network assortativeness. Therefore, summing local assortativeness values of all individual nodes should result in the network assortativeness. Two networks with the same network assortativeness and similar degree distributions may have entirely different *local assortativeness distributions*, or profiles. Thus, local assortativeness profiles can give a new perspective about the design features of a network. For example, we could see whether the largest hubs in a network are mostly connected to each other, or to smaller hubs and peripheral nodes. Interconnected giant hubs may be a sign of robustness against targeted attacks. As another example, the local assortativeness distribution may indicate whether all peripheral nodes are connected to more hub-like nodes (*i.e.*, indicating presence of *star* motifs), or to one another (indicating *chain* motifs). We analyze simulated and real-world scale-free networks (*e.g.*, biological networks) based on their local assortativeness profiles, and highlight motif and design features within them. This shows that the local assortativeness distribution is related to network's robustness against

(a) E-mail: mikhail.prokopenko@csiro.au

targeted attacks, and that node roles can be classified based on their local assortativeness values.

Definitions and terminology. – To formally define the concept of local assortativeness, we will first introduce a number of related definitions. For simplicity, we introduce these definitions for networks with undirected links, which can then easily be extended to directed networks. The degree distribution $p(k)$ is a probabilistic distribution of encountering a node with a given number of links [6]. A related concept is the remaining degree distribution $q(k)$, where the *remaining degree* means the number of 'remaining links' of a node that one sees when one travels along a randomly chosen link toward the node [1]. It can be shown that compared to the degree distribution p, the remaining degree distribution q is biased towards higher degrees, and the relationship between the two distributions is given by

$$q(k) = \frac{(k+1)p(k+1)}{\sum jp_j}. \tag{1}$$

Given $q(k)$, one can introduce the quantity $e_{j,k}$ as the joint probability distribution of the remaining degrees of the two nodes at either end of a randomly chosen link. Given these quantities, network assortativeness is defined [2,3] as a correlation function which is zero for non-assortative mixing, and positive or negative for assortative or disassortative mixing, respectively:

$$r = \frac{1}{\sigma_q^2}\left[\sum_{jk} jk\left(e_{j,k} - q(j)q(k)\right)\right], \tag{2}$$

where σ_q is the standard deviation of the remaining degree distribution of the network, $q(k)$. Similarly, the term $\sum_j jq(j)$ can be understood as μ_q, the expected value or mean of the remaining degree distribution. Therefore network assortativeness r can be defined as

$$r = \frac{1}{\sigma_q^2}\left[\left(\sum_{jk} jke_{j,k}\right) - \mu_q^2\right], \tag{3}$$

where μ_q and σ_q are both constants for the network. Having this refined but essentially unchanged definition of network assortativeness, let us now concentrate on defining local assortativeness of a node.

We propose to define *local assortativeness* as the contribution of a given node to the network assortativeness, which means we need to determine how much each node contributes to the term $\frac{1}{\sigma_q^2}[(\sum_{jk} jke_{j,k}) - \mu_q^2]$. Let us first look at the term $\sum_{jk} jke_{j,k}$ and the contribution of each node to this term. Suppose we visit all the nodes in a network, and from each node in turn we visit all the links of that node. In a network with N nodes and M links, the total visits we will thus make will be $2M$, since each link will be visited twice, once from each end. Suppose we build up the probability distribution e_{jk} as we make these visits. Each link will add a probability of $(1/2M)$ to the pair of (j,k), where j and k are the remaining degrees of nodes at each end of the link. Thus, each visit to a link will contribute $jk/2M$ to the sum $\sum_{jk} jke_{j,k}$. Therefore, considering a node with remaining degree j (fig. 1) which is connected to nodes with remaining degrees k_1, $k_2, \ldots k_{j+1}$, it will contribute $(jk_1/2M) + (jk_2/2M) + \cdots + (jk_{j+1}/2M) = \frac{j}{2M}\sum_{i=1}^{j+1} k_i$ to the sum $\sum_{jk} jke_{j,k}$. Let us denote the average remaining degree of a node's neighbours as $\overline{k} = \frac{1}{j+1}\sum_{i=1}^{j+1} k_i$. Then the individual node's contribution, α, to the sum $\sum_{jk} jke_{j,k}$ is

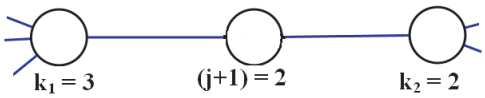

Fig. 1: The considered node has one remaining degree ($j = 1$), *i.e.*, it has two neighbours: n_1 and n_2 with three and two remaining degrees respectively ($k_1 = 3$ and $k_2 = 2$). The average remaining degree of the neighbours $\overline{k} = 2.5$.

$$\alpha = \frac{j}{2M}\sum_{i=1}^{j+1} k_i = (j+1)\frac{j\overline{k}}{2M}. \tag{4}$$

Now let us consider a node's contribution to the term μ_q^2. Distribution $q(k)$ is defined in terms of traversals along links. Therefore a node's contribution to the mean of $q(k)$ is proportional to the number of links it has, and it has to be scaled by $(j+1)/2M$, where $(j+1)$ is the node degree. Thus we define a node's contribution β to μ_q^2 as

$$\beta = (j+1)\frac{\mu_q^2}{2M}. \tag{5}$$

The standard deviation is already a scaling term, and we need not worry about a single nodes' contribution to it. Consequently, we formally define local assortativeness ρ as

$$\rho = \frac{\alpha - \beta}{\sigma_q^2} = \frac{(j+1)\left(j\overline{k} - \mu_q^2\right)}{2M\sigma_q^2}, \tag{6}$$

where j is the node's remaining degree, $\overline{k}$ is the average remaining degree of its neighbours, and $\sigma_q \neq 0$. By including the scaling term, we ensure that the equation for local assortativeness satisfies the condition

$$r = \sum_{i=1}^{N} \rho_i. \tag{7}$$

If the standard deviation σ_q is zero, then the definition (6) cannot be applied. This case, however, can arise only if the network is homogeneous, *i.e* all nodes in the network have the same degree. Since local assortativeness is defined as the contribution of each node to the network assortativeness, we state that in this case all contributions are equal, and if $\sigma_q = 0$, the local assortativeness is set to $\rho = 1/N$.

Since local assortativeness is a property of a node, it is possible to construct local assortativeness distributions

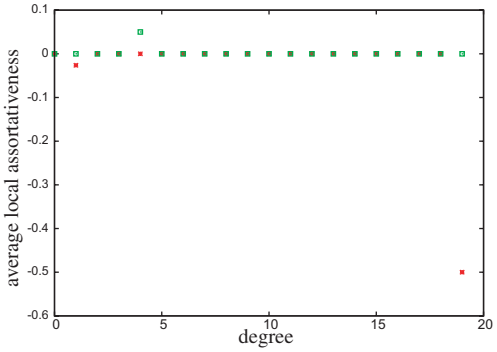

Fig. 2: Local assortativeness distribution, $\overline{\rho}(k)$ *vs.* k, of a regular lattice with four nodes connecting to each node (squares), and of a star graph (stars). Network size in both cases is $N = 20$.

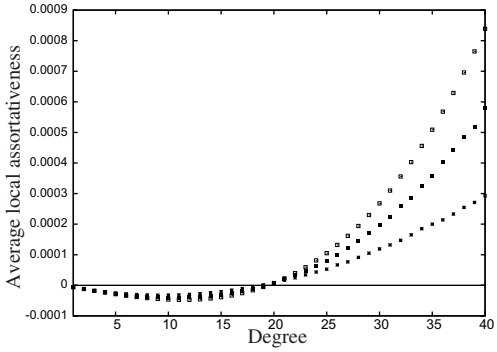

Fig. 3: Local assortativeness distributions for assortative networks ($r \approx 1$, squares; $r \approx 0.5$, filled squares) and non-assortative networks ($r \approx 0$, stars): $\overline{\rho}(k)$ *vs.* k.

for a given network, plotting local assortativeness values against degrees. Since nodes with the same degree can have various local assortativeness values, we could represent either the total of the local assortativeness values of nodes with a given degree k, or the average local assortativeness value for all nodes with a given degree k. In this letter, we calculate the average value of local assortativeness for all the nodes in a network with a given degree k, denoting this value as $\overline{\rho}(k)$. Therefore, if we denote by $N(k)$ the number of nodes with degree k, then the network assortativeness is

$$r = \sum_k N(k)\overline{\rho}(k), \tag{8}$$

where $\overline{\rho}(k)$ is the average local assortativeness of all nodes with degree k. This equation can also be written as

$$r = N \sum_k p(k)\overline{\rho}(k), \tag{9}$$

where $p(k)$ represents the degree distribution of the network. Henceforth, we shall mostly use the "average local assortativeness *vs.* degree", $\overline{\rho}(k)$ *vs.* k, distributions.

Now let us look at the local assortativeness distributions of some classical network structures before analyzing real world networks and their local assortativeness.

Local assortativeness in model networks. –

Regular lattice. Lattice-like networks are common in some human-designed architectures, particularly parallel computers [1]. For a lattice-like network each node has the same degree and remaining degree, therefore the variance of remaining degree distribution is 0. Since there is only one type of nodes, the network is perfectly assortative ($r = 1$) and the local assortativeness of all nodes is $1/N$. Therefore the local assortativeness distribution resembles Kronecker delta-function (fig. 2).

Star network. Star graph is another common motif, responsible for the small world feature in many networks. In a pure star graph, any given link has a peripheral node at one end, with degree one (*i.e.* remaining degree zero). It can be shown that a star graph is perfectly disassortative ($r = -1$). Furthermore, any node in the star graph has either its remaining degree as zero, or all of its neighbors' remaining degrees as zero. Thus, a hub does not always have to have positive local assortativeness, as the latter depends on both the node degree and neighbours degrees. It is easy to see that the term represented by eq. (4) reduces to zero in all cases. Thus the local assortativeness reduces to

$$\rho = -\frac{j+1}{2M}\frac{\mu_q^2}{\sigma_q^2}. \tag{10}$$

Figure 2 shows the local assortativeness distribution for a pure star graph: the central node is much more locally-disassortative, as it connects with many dissimilar nodes, whereas the peripheral nodes are less locally-disassortative since they connect to only one dissimilar node.

Local assortativeness in scale-free networks. – Now we proceed to analyze the local assortativeness distributions for simulated and real world scale-free networks. Most real world networks, including biological networks, social networks and technological networks including Internet and World Wide Web are scale-free networks with power law degree distributions [6–8]. The degree distributions of such networks take the form of $p(k) = k^{-\gamma}u(\frac{k}{N_p})$, where u is a step function which specifies a cut-off at $k = N_p$. Hence we focus mainly on scale-free networks and their local assortativeness profiles. First of all, we study simulated scale-free networks with various assortativeness values (constructing them using a method described in [3]). Some typical results are shown in figs. 3 and 4. The local assortativeness for all nodes (ρ_i *vs.* i) is shown as a scatter plot in figs. 5, 6, and 7.

A number of interesting facts can be observed from these distributions. First of all, there is a large number of disassortative nodes, $\overline{\rho} < 0$, in any network, regardless of whether the network is overall assortative, disassortative, or non assortative. Conversely, the number of assortative nodes, $\overline{\rho} > 0$, in a disassortative network is very small, or even zero. Secondly, a large proportion of nodes in any network fall on the "slightly disassortative" area (just below the zero axes in the scatter plots —see

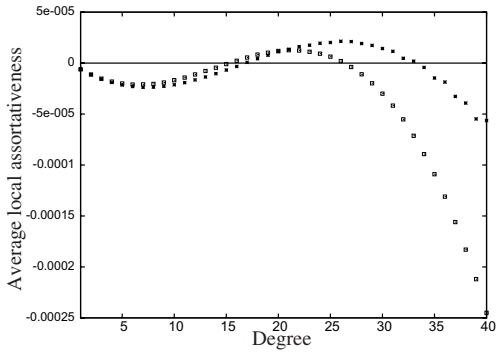

Fig. 4: Local assortativeness distribution for disassortative networks ($r \approx -1$, squares; $r \approx -0.5$, stars): $\overline{\rho}(k)$ *vs.* k.

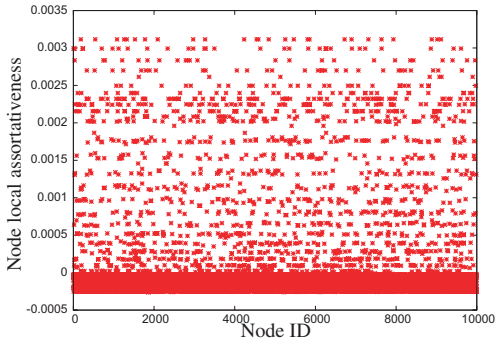

Fig. 5: Local assortativeness distribution for assortative networks ($r \approx -1$): ρ_i *vs.* i. The top layer(s) contain hubs.

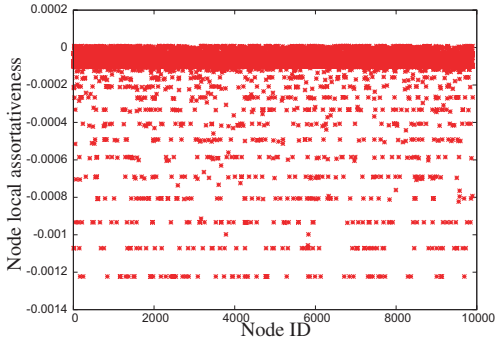

Fig. 6: Local assortativeness distribution for disassortative networks ($r \approx -1$): ρ_i *vs.* i. The bottom layer(s) contain hubs.

figs. 5, 6, and 7). These correspond to peripheral nodes, which must exist in any type of scale-free network. The hubs however, interestingly, have either high local assortativeness or high local disassortativeness. Specifically, the hubs seem to be highly assortative in assortative ($r \approx 1$) or non-assortative ($r \approx 0$) networks, whereas the hubs in disassortative networks ($r \approx -1$) are highly disassortative themselves. Thus the assortativeness or disassortativeness of the hubs seem to determine the assortativeness or disassortativeness of the network to a large extent, even though hubs are much smaller in number than peripheral nodes. Finally, we observe that when the network is very assortative or very disassortative, the nodes can be more easily clustered together based on local assortativeness (the horizontal stratification into levels of nodes, observed in

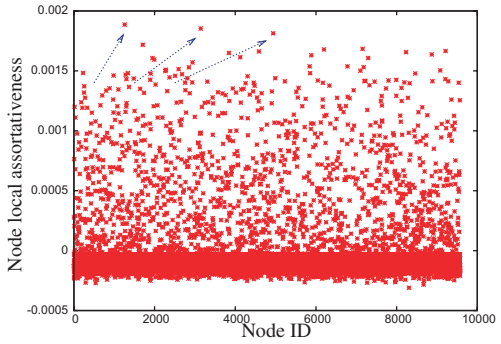

Fig. 7: Local assortativeness distribution for non-assortative networks ($r \approx 0$): ρ_i *vs.* i. Giant hubs are marked by arrows.

figs. 5 and 6). There are clear "levels" of hubs of a similar local assortativeness, as well as provincial hubs and inner peripheral nodes. Thus, local assortativeness can be used to classify nodes based on their function in these cases. In contrast, when the network is non-assortative ($r \approx 0$), nodes cannot be easily grouped based on their local assortativeness. There is no stratification into levels (clusters) clearly visible in fig. 7. Furthermore, a small number of giant hubs are likely to emerge in non-assortative networks (see fig. 7). When the network is more assortative or more disassortative, a larger number of smaller hubs replace this small group of giant hubs. There are no corresponding giant hubs in these cases.

Why is it the case that the giant hubs are more likely to emerge in relatively non-assortative networks? In assortative networks, hubs tend to connect to each other (this must happen for high assortativeness), preventing a single giant hub from emerging. In disassortative networks, there are many smaller isolated hubs which are not connected to each other (many star motifs). If these hubs make connections between each other or other provincial hubs the disassortative nature begins to be compromised. It is therefore in non-assortative networks that single giant hubs tend to emerge. This implies that non-assortative scale-free networks ($r \approx 0$) are more vulnerable to targeted attacks which destroy hubs.

Local assortativeness in real-world networks. – We have also studied local assortativeness distributions for a number of real-world networks which are scale-free, focusing mainly on biological networks and comparing them with human-designed networks. Our study included a broad range of networks [9,10], including specifically metabolic networks, food webs, citation networks, and Internet.

It can be noted that most of the networks we studied are disassortative in nature. However, the question is, are all the nodes in them disassortative? In other words, do they connect to nodes that are always dissimilar to them in degrees? We found that there are fundamental differences in this regard between the various studied networks.

We can observe that the Internet, fig. 8, has only locally disassortative nodes —that is, nodes connecting

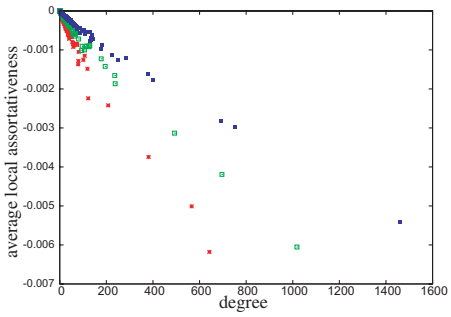

Fig. 8: $\overline{\rho}(k)$ *vs.* k, of Internet Autonomous System level 1998 ($r = -0.198$, stars); 1999 ($r = -0.174$, squares); 2000 ($r = -0.16$, filled squares).

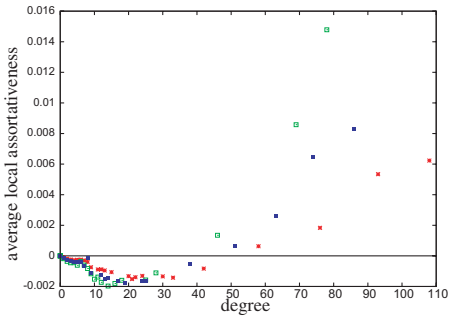

Fig. 9: $\overline{\rho}(k)$ *vs.* k, of metabolic networks A. thaliana ($r = -0.171$, stars); B. burgdorferi ($r = -0.16$, squares); A. pernix ($r = -0.182$, filled squares).

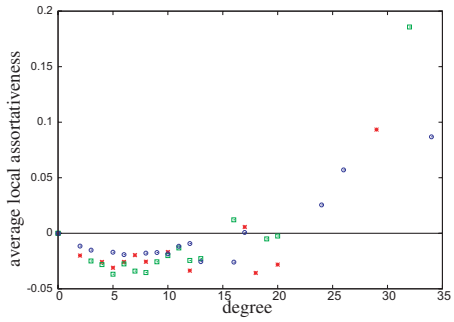

Fig. 10: $\overline{\rho}(k)$ *vs.* k, of food webs Crystal River Creek D ($r = -0.467$, stars); Crystal River Creek C ($r = -0.334$, squares); Lower Chesapeake Bay in Summer ($r = -0.391$, filled squares).

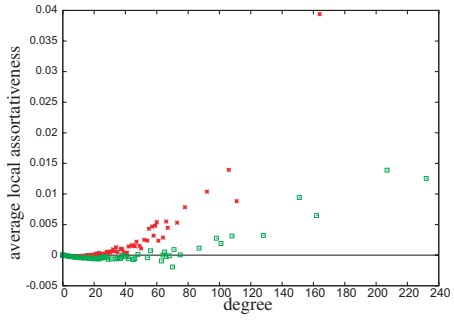

Fig. 11: $\overline{\rho}(k)$ *vs.* k, of Citations Scientometrics ($r = -0.03$, stars); Small & Griffith and Descendants ($r = -0.193$, squares).

to other nodes that are dissimilar in degrees. Moreover, the local assortativeness decreases almost linearly with node degree— that is, the hubs are the ones that mainly contribute to network disassortativeness. That must mean that the hubs are not often connected to other hubs, but connected to peripheral nodes.

In metabolic networks, fig. 9, however, we encounter a different scenario. While both the metabolic networks and Internet are overall disassortative, their local assortativeness profiles are quite contrasting, especially in the region of higher degrees. In metabolic networks, there are some locally highly assortative nodes, even though the whole network is disassortative. Moreover, these assortative nodes are hubs. The local assortativeness curve decreases until a point in the vicinity of smaller hubs ($k \approx 20$), which are most disassortative. Then the curve starts rising again, leading to higher-degreed nodes that are locally very assortative. So we can infer that in these networks the largest hubs have a lot of links among themselves; the medium-sized hubs ($k \approx 60$) connect more or less randomly vis-à-vis node degree, whereas the smaller hubs ($k \approx 20$) connect to peripheral nodes, giving the network the overall disassortativeness. The main difference, when compared with the Internet, is that the decrease in local assortativeness reverses in metabolic networks when we encounter smaller hubs.

The food webs, fig. 10, and citation networks, fig. 11, display similar profiles to metabolic networks, albeit somewhat less obviously. The food webs are much smaller in size ($N \approx 100$ as opposed to thousands of nodes in metabolic networks) preventing a smooth curve in the local assortativeness profile, although a transition point is evident too (*e.g.*, $k \approx 15$ in Chesapeake bay food web). Both figures show that the hubs are locally highly assortative, analogously to metabolic networks and unlike the Internet. Citation networks manifest another feature —the decrease in local assortativeness of less-connected nodes is less pronounced, but still is replaced by the reverse trend toward higher local assortativeness ($k \approx 50$).

The observed profiles can be explained if one considers the growth process of the respective networks in terms of evolution (natural or technological). The growth process can be roughly separated into two parts: initial phase, and mature phase. The studied network types have some similarities and differences with respect to these phases. The metabolic networks experienced a relatively vigorous initial phase during some evolutionary past: new biochemical substrates were added to the mix, new links were actively formed. As the biochemistry matured, this process became much more constrained (fine-tuning), as new viable combinations became scarce. Hence, we observe the difference between small hubs, whose connections formed mostly during the initial phase, providing network fragments, and the large hubs that acquired connections during the mature phase.

A similar tendency is evident in food webs: a relatively active initial phase during some evolutionary past when species established their local niches (forming fragments

of the network), and a mature phase, when food-webs filled the whole relevant environment, fine-tuning the relationships. The citation networks (exemplifying social networks) show that the initial phase of the growth process is relatively less vigorous (the decrease in local assortativeness is less pronounced), but still leads to a mature phase, when some authors have established a lot of collaborators. Interestingly, such a two-phase interpretation informs that the Internet has not yet achieved a mature phase (the decrease in local assortativeness does not reverse), and may acquire some locally-assortative nodes in the future. In short, the transition in the local assortativeness profile is a good indicator of growth history, pinpointing qualitatively different phases.

Another possible explanation is that Internet was designed, or evolved in a comparatively much quicker time, with relatively unchanging sets of goals. The metabolic networks, on the other hand, have evolved quite gradually, with a set of constraints and goals that changed as the network evolved [11]. This is reflected in the patterns enabling the network to adapt, with minimal changes in its connectivity, if the overall network functionality is required to change. It has been suggested that adding or deleting a node in a network is much more costly than adding or deleting a link [11]. Biological networks, which evolved through changing constraints and goals, must be able to reconfigure themselves with minimum cost when needed. It is likely that having hubs with high local assortativeness (*i.e.* having the hubs closely connected together) helps to achieve this. This is a design feature that human-designed networks will have to adopt, if they are to be robust to changing performance goals.

We would like to contrast local assortativeness ρ, with the rich-club connectivity ϕ defined as an average connectivity of nodes that have more than a specified number of degrees [12,13]. While the latter is computed over sub graphs (rich clubs), the former is a measure of a specific node. In particular, it is possible to measure ρ for any peripheral node, but it is not possible to compute ϕ for peripheral nodes alone, as they would have to belong to a rich club. One may consider a cumulative average local assortativeness, $\mathcal{R}(k)$, by aggregating $\rho(k)$ for all degrees higher than k. Contrasting $\mathcal{R}(k)$ with the corresponding ϕ reveals that these quantities are correlated, but the correlation is non linear. A more detailed comparison between these measures is a subject of future research.

Conclusion. – Many real-world networks, including biological and technical networks, display degree assortativeness or disassortativeness, where nodes preferentially connect to each other based on their degrees. There can be various motifs or subnetworks within the overall network, whose assortativeness might be markedly different from the overall network assortativeness. In other words, various nodes in a network contribute in different ways to network assortativeness, and in some cases, these nodes may display assortative tendencies even though the network is overall disassortative, or vice versa.

We formally defined a measure of local assortativeness to quantify a node's contribution to network's overall assortativeness, and studied local assortativeness profiles for both simulated and real networks. We pointed out that any scale-free network must have a large number of locally disassortative nodes in it, but it may not have any locally assortative nodes at all. The non-assortative networks were shown to be more likely to have giant hubs, and therefore to be most vulnerable to targeted attacks. In practical terms, local assortativeness can be used to cluster nodes based on their relative importance when faced with targeted attacks against the network.

We compared biological, social and technological networks by their local assortativeness profiles, and found significant differences. Specifically, the biological networks have hubs with high local assortativeness, regardless of the overall assortativeness of the network. This feature indicates a phase of growth maturation, strengthening the robustness of biological networks against changing functional goals. Local assortativeness can be defined for subnetworks, motifs, or network modules. A precise relation between the network assortativeness and the assortative patterns of these modules is a subject of future research.

REFERENCES

[1] Solé R. V. and Valverde S., *Information theory of complex networks: on evolution and architectural constraints,* in *Complex Networks*, edited by Ben-Naim E., Frauenfelder H. and Toroczkai Z., *Lect. Notes Phys.*, Vol. **650** (Springer) 2004.

[2] Newman M. E., *Phys. Rev. Lett.*, **89** (2002) 208701.

[3] Piraveenan M., Prokopenko M. and Zomaya A., *Information-cloning of scale-free networks,* in *Advances in Artificial Life: 9th European Conference on Artificial Life (ECAL-2007)*, edited by e Costa F. A., Rocha L. M., Costa E. and I. Harvey A. C., *Lect. Notes Artif. Intell.*, Vol. **4648** (Springer) 2007.

[4] Newman M. E., *Phys. Rev. E*, **67** (2003) 026126.

[5] Callaway D. S., Hopcroft J. E., Kleinberg J. M., Newman M. E. and Strogatz S. H., *Phys. Rev. E*, **64** (2001) 041902.

[6] Albert T. R. and Barabasi A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.

[7] Albert R., Jeong H. and Barabasi A.-L., *Nature*, **401** (1999) 130.

[8] Faloutsos M., Faloutsos P. and Faloutsos C., *Comp. Commun. Rev.*, **29** (1999) 251.

[9] Centre for Complex Network Research University of Notre Dame, `http://www.nd.edu/~networks/` (2008).

[10] Batagelj V. and Mrvar A., *Pajek datasets*, `http://vlado.fmf.uni-lj.si/pub/networks/data/` (2006).

[11] Alon U., *Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall, London) 2007.

[12] Zhou S. and Mondragón R. J., *IEEE Commun. Lett.*, **8** (2004) 180.

[13] Colizza V., Flammini A., Serrano M. A. and Vespignani A., *Nat. Phys.*, **2** (2006) 110.