# Evaluating the interplay between different nodes characteristics in complex networks

Matteo Cinelli, Giovanna Ferraro, and Antonio Iovanella

Department of Enterprise Engineering, University of Rome Tor Vergata,
Via del Politecnico, 1 - 00133, Rome Italy,
matteo.cinelli@uniroma2.it, giovanna.ferraro@uniroma2.it,
antonio.iovanella@uniroma2.it

In complex networks literature the idea that nodes can possess characteristics able to shed light on their role in the system has been widely accepted and investigated [1, 2, 5, 7]. Thus the possibility to assess the relevance of nodes taking advantage of their properties, either provided by the system or assigned using external arguments, represents a topic of interest within many applications [2, 5, 6, 7]. This kind of approach has been used in order to study homophilic trends in networks by Park and Barabási [7] through the observation of a phenomenon called *dyadic effect*. The quantification of such effect has been performed by two measures called *dyadicity* and *heterophilicity* that are defined using the type of dyads that connect similar nodes. Considering nodes characterized with a binary feature $c_i$ we encounter then three kind of dyads called $m_{11}$, $m_{10}$ and $m_{00}$. Using these arguments, dyadicity has been defined as $D = \frac{m_{11}}{\overline{m}_{11}}$ while heterophilicity as $H = \frac{m_{10}}{\overline{m}_{10}}$ where, given the number $n_1$ of nodes with $c_i = 1$, $\overline{m}_{11}$ and $\overline{m}_{10}$ are the expected values of $m_{11}$ and $m_{10}$ respectively. Dyadicity and heterophilicity consequently define a two dimensional space called $D-H$ space that constitutes a region where to investigate the distribution of nodes characteristics when expressed by binary values. Then, if provided with a set of nodes characteristics, they can be analyzed one at the time, computing for each one: the deviation of its distribution from random and the correlation with the network structure using the values of $D$ and $H$. These two measures let us consequently gain some important insights about the meaningfulness of a property shared by a certain number $n_1 \in N$, where $N$ is the set of nodes. The described methodology is currently applied using complete enumerations or heuristic algorithms and visually explained by a tool called phase diagram that shows the distribution of nodes characteristics within the $D-H$ space. This approach has been adopted thanks to its peculiarity in bringing together some endogenous elements, related to the topology of the network, with some other exogenous elements related to the nodes characterization but only applied to very small networks because of its hardness that grows exponentially with the network size. Despite these aspects, the described methodology becomes of particular interest when we have to deal with many nodes characteristics and we intend to make valid hypothesis on which one has been more relevant in determining the actual network topology. Under these circumstances our approach is devoted to provide an informative content which represents a reasonable alternative to the one

COMPLEX NETWORKS

The $5^{th}$ International Workshop on Complex Networks & Their Applications. Nov. 30 - Dec. 02, 2016, Milan (Italy)

of a complete enumeration, in a less onerous way, by tackling the complexity of the problem through a new methodology that doesn't require the computation of the phase diagram.

Based on certain analytical arguments deriving from structural bounds on the existence of dyads type [7] we compute, for any $n_1 \in N$ and for any possible distribution of a node characteristic, two curves: the first using the values of $m_{11max} = \binom{n_1}{2}$ and $m_{10max} = n_1(N - n_1)$; the second using $\overline{m}_{11} = \binom{n_1}{2}d$, $\overline{m}_{10} = n_1(N - n_1)d$ where $d$ is the network density. Moreover we compute $D_{max}, H_{max}$ comparing them with the value related to the currently examined configuration and introducing a novel index of Relevance. This kind of approach let us directly discern, among many characteristics, the ones that can be considered as most important and that can eventually deserve further analyses using, for instance, the methods described in [7]. Since different values of $D$ and $H$ define a set of trends, called dyadic if $D > 1$, anti-dyadic if $D < 1$, heterophilic if $H > 1$ and heterophobic if $H < 1$, that can be expressed as a percentage if compared to their maximum and minimum values, the index of Relevance can be computed as the arithmetic mean of two components called $R_D$ and $R_H$ (Table 1) that, fixed $n_1$, represent relative values of $D$ and $H$. The proposed framework can be implemented in many

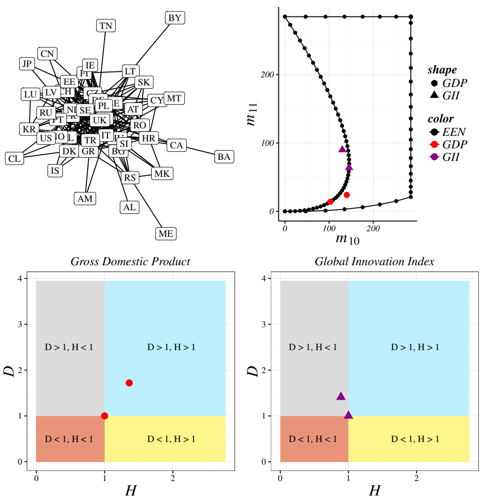


Fig. 1: Clockwise representation of: the EEN members countries in 2011; Upper bound and expected values of $m_{11}$ and $m_{10}$ curves for each $n_1 \in N$ and coordinates of the GDP and GII points for $n_1 = (11, 23)$; Relevance components ($R_D = 0.25$; $R_H = 0.2$; $R = 0.225$) of the GDP in the $D - H$ space given $n_1 = 11$; Relevance ($R_D = 0.14$; $R_H = 0.11; R = 0.125$) of the GII in the $D - H$ space given $n_1 = 23$.

domains of application and becomes valuable when we have to deal with interactions among countries whose performances are measured using a wide range of indicators.

We apply this methodology to a real case study [3, 4] over three years, an initiative financed by the European Commission called Enterprise Europe Network (EEN), in which nodes are the member countries, links are partnership agreements of technological transfer among them. We consider as nodes characteristics a set of macroeconomic indicators such as the Gross Domestic Product (GDP) and the Global Innovation Index (GII). We evaluate how countries that over-perform ($c_i = 1$) and under-perform ($c_i = 0$) considering the mean value of each index, from 2011 to 2013, are connected among themselves as in [5] and how this phenomenon affected the network structure. An example of application is reported in Figure 1 where we observe how the GDP, which shows a higher Relevance, is more significant than the GII in the rationale behind the drafting of agreements.

Table 1: Values of $R_D$ and $R_H$

| | $R_D$ | | $R_H$ |
|---|---|---|---|
| $D > 1$ | $1 - \frac{D_{max} - D}{D_{max} - 1}$ | $H > 1$ | $1 - \frac{H_{max} - H}{H_{max} - 1}$ |
| $D < 1$ | $1 - D$ | $H < 1$ | $1 - H$ |

## References


1. Bianconi, G., Pin, P., Marsili, M.: Assessing the relevance of node features for network structure. Proceedings of the National Academy of Sciences 106(28), 11433–11438 (2009), `http://www.pnas.org/content/106/28/11433.abstract`
2. Cinelli, M., Ferraro, G., Iovanella, A.: Some insights into the relevance of nodes' characteristics in complex network structures. In: Zylka, M.P., Fuehres, H., Colladon, A.F., Gloor, P.A. (eds.) Designing Networks for Innovation and Improvisation, Proceedings of the 6th International COINs Conference. Springer Proceedings in Complexity, Springer International Publishing (2016)
3. Ferraro, G., Iovanella, A.: Organizing collaboration in inter-organizational innovation networks, from orchestration to choreography. International Journal of Engineering Business Management 7(24) (2015)
4. Ferraro, G., Iovanella, A.: Revealing correlations between structure and innovation attitude in inter-organisational innovation networks. International Journal of Computational Economics and Econometrics 6(1), 93–113 (2016)
5. Ferraro, G., Iovanella, A., Pratesi, G.: On the influence of nodes' characteristic in inter-organisational innovation networks structure. International Journal of Computational Economics and Econometrics 6(3), 239–257 (2016)
6. Hu, T., Andrew, A.S., Karagas, M.R., Moore, J.H.: Functional dyadicity and heterophilicity of gene-gene interactions in statistical epistasis networks. BioData mining 8(1), 1 (2015)
7. Park, J., Barabási, A.L.: Distribution of node characteristics in complex networks. Proceedings of the National Academy of Sciences 104(46), 17916–17920 (2007)