A Quantitative Theory of Social Cohesion

From Sociology to Network Science to Psychology

Adrien Friggeri Eric Fleury
Inria / ENS de Lyon
France

With the democratization of computing power and access to rich datasets in the past couple of decades, sociology has witnessed a shift towards a data-oriented and network-based vision of society. More and more, sociological studies move away from basic metrics attached to individuals such as degree, clustering and centrality and focus on the larger patterns of the network such as social communities. Here, we will first introduce a formal model of the community, which contrasts with existing models both in that it provides a mathematical description of the community, whereas several other proposals describe a partition of the network in communities. Moreover, it was validated and confronted to reality through a large scale experiment in which it was observed that it correlated strongly to the subjects' perception of their communities. In a second time, we will leverage this higher level topological description of social network to exhibit links between the underlying social network of subjects and their personality. In effect, this work bridges Sociology and Psychology through the use of Network Science.

In 1957, Hillery was interested in understanding what people meant by "community". Using a corpus of 98 different definitions of the term, he found out that the only feature which was common to all of those definitions was that "community" was somehow related to people – it must however be noted that a majority of those definitions also included a notion of "something shared" between the members of the community. This observation was then repeated by Poplin in 1979 and Stuckey in 2007, which invariably lead to the same conclusion: the notion of community was a blurry one. Nevertheless, in the field of social network analysis, although there was no consensus on what a community was, there have been a large number of proposals to achieve the effect of finding or extracting communities from networks. The study of communities in social networks has evolved along two orthogonal directions. On the one hand, in the tradition of classical blockmodelling, there has been a trend in dividing a social network into disjoint communities – heralded by the works of Newman et al. (2004) – which has in the recent years revolved around the modularity, a quality function which rates a partition with respect to the underlying network. On the other hand, the realization that partitioning is too drastic to precisely describe social communities – given that one individual should be able to be part of several social groups, such as his family and his group of friends – there has been an effort geared towards the analysis of overlapping communities. However, since it has been repeatedly observed that there was no consensus on what a community is, these work have mainly been constructed upon unverified assumptions on the notion of community. Intuitively, all converge to the idea that community is a dense set of people, relatively isolated from the rest of the network – the formal definitions of "dense" and "isolated" varying wildly from one author to the other.

We have proposed a metric, called the *cohesion*, which not only embraces these assumptions but was also constructed using well established sociological concepts and was then verified through a large scale experiment. At its most basic level, the *cohesion* rates the communitiness of a set of nodes, independently from the rest of the network and the existence of other communities. Among other things, it builds upon the notion of triadic closure, a property which states that, given three nodes A, B and C, if

there is a strong tie between A-B, and A-C, then there is at least a weak tie between B-C. Although this notion is too strong to apply to large and complex networks, it has usually been accepted as a useful simplification. It can however be loosen to apply more broadly: instead of considering ties of identical nature, one can suppose the existence of different type of ties – family, work, friendship, etc. – and the principle can be reformulated as *given A,B and C, if there is a strong tie of type T between A-B and A-C, then there is at least a weak tie of type T between A-C,* this generalization applies more broadly to real world complex social networks. Building upon this generalized notion, we construct the *cohesion*, which rates the density and isolation of a community in terms of the existence of closed triads. That is, the cohesion[1] of a set of *k* nodes is defined as:

$$\mathcal{C} = \frac{\triangle}{\binom{k}{3}} \times \frac{\triangle}{\triangle + \text{⊂}}$$

Where $\triangle$ represents the number of closed triads included in the community, ⊂ the number of closed triads containing exactly two nodes belonging to the community The cohesion is as such defined as a product of a triangular density or transitivity (left factor) and isolation (right factor).

The use of the cohesion was validated through a large scale experiment on Facebook called Fellows[2]. More than 3000 subjects connected there Facebook account and shared the network of their friends – the interconnections between their friends. Leveraging this information and using a simple greedy algorithm to maximize the cohesion, we computed a covering of their social neighborhood in communities. It must be noted that the greedy algorithm is far from being optimal, which purposefully lead to the generation of communities with high as well as low cohesion. The subjects were then asked to give a rating between 1 and 4 to each of those communities. We then confronted those ratings to the cohesion (Figure 1.) and observed that both quantities were highly correlated – we observed a high Spearman rank correlation between those two values ($\rho = 0.91$, p-value $= 4.3 \times 10^{-38}$), as well as a high linear Pearson correlation between the logarithms of these values ($r = 0.97$, p-value $= 3.2 \times 10^{-63}$). From this, given that there are strong links between the metric and the subjective perception of the participants, we conclude that the cohesion is a good indicator of the communitiness of a set of nodes.

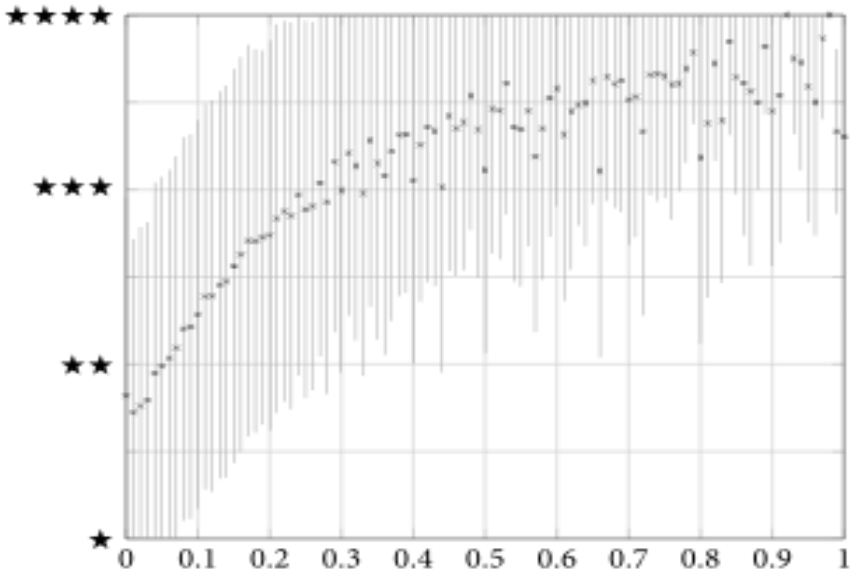

Figure 1. Average rating as a function of the cohesion.

[1] A. Friggeri, G. Chelius, and E. Fleury, "Triangles to Capture Social Cohesion," presented at the 2011 IEEE Third International Conference on Social Computing (SocialCom 2011), Cambridge, United States

[2] http://fellows-exp.com

Having a quantitative model of community, which fits closely what members of communities perceive, allows us to describe more precisely the topological features of the network. Rather than being limited to degree, clustering, centrality or other node-related metrics, the algorithmic optimization of the cohesion would allow the discovery of a community profile for each node – that is, a description of a node's communities, and their interactions with one another. It should be noted here that we have shown that problem of finding an optimally cohesive community containing a prescribed set of nodes is NP-Hard. We have however proposed C^3, a heuristic[3] that computes a covering of a network in cohesive – potentially overlapping – communities, using a strategy of optimizing the cohesion and the transitivity independently from one another, and has provided successful results.

Sociology in general, and social network analysis in particular, have often had a mean field approach to the complexity and heterogeneity of individuals characteristics. Although the links between basic network metrics – eg. degree – and personality traits have been studied, the correlations between the deeper social structure and the psychology of the actors have been relatively ignored up to this date. Using two datasets, one containing psychological data from the myPersonality website – a service which provides online psychological tests to Facebook users – and the sociological information pertaining to those users extracted from Facebook – the set of the user's friends, and the edges between those friends – we were able to show deep links between the community structure of the subjects social neighborhood and their personality traits[4]. Specifically, we have focused on the five-factor model of personality – Openness, Conscientiousness, Agreeableness, Extraversion, Neuroticism – and of those five we have only found significant correlations between descriptors of the social topology and Extraversion. It is already well know that the trait of extraversion correlates with the degree of a subject. We have however observed that:

- Extraversion correlates positively with the number of cohesive communities a subject belongs to, which is linked to the previous result with the degree;
- Extraversion correlates negatively with the average sizes of communities a subject belongs to, which we interpret as a sign that introverts tend to hide in larger communities in order to attract less attention;
- Extraversion correlates positively with the variability of the cohesion of communities a subject belongs to, which we understand in terms of social heterogeneity and adaptability: extroverts have a tendency of being part of both tight and loose communities;
- Extraversion correlates positively with the amount of overlap in the division into communities of the subject's neighborhood – as we have stated earlier communities are not disjoint, and one way of quantifying the extent to which those communities are woven one into another is to observe the fraction of nodes which are present in more than one community. This observation is compatible with the hypothesis that extroverts act as bridges between communities and introduce individuals form one community to those in another one, whereas introverts tend to shy away from unusual social interactions.

In conclusion, the notion of social communities as defined in terms of *cohesion* is not only socially valid as exhibited by the Fellows experiment, but also appears to be connected to the psychology of the actors in a network. As such, it acts as a bridge between network science, sociology and psychology. Future works are largely geared towards the study of ties between a subject's social influence and the community structure of their neighborhood.

[3] A. Friggeri and E. Fleury, "Finding cohesive communities with C3," Technical Report, INRIA, RR-7947, Apr. 2012.
[4] A. Friggeri, R. Lambiotte, M. Kosinski, and E. Fleury, "Psychological Aspects of Social Communities," presented at the 2012 ASE/IEEE International Conference on Social Computing (SocialCom 2012), Amsterdam, Netherlands.