

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A Mesoscale Description of Networks' Dynamics Through Continuous Partitioning

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1651339> since 2017-11-06T19:16:09Z

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

A mesoscale description of networks' dynamics through continuous partitioning

Keywords: networks, dynamics, mesoscale, science history, continuous

Introduction: Large graphs require effective methods giving an appropriate mesoscopic description. Several approaches exist today to partition (static) graphs into communities (Fortunato, 2010). However, many networks are intrinsically dynamical, and describing them as static networks would cause loss of important information (Holme and Saramaki, 2012 ; Holme 2015). For example, dynamic processes such as the emergence of new scientific disciplines, their fusion, split or death need a mesoscopic description of the evolving network of scientific articles.

There are two straightforward approaches to describe an evolving network using methods developed for static networks. The first finds the community structure of the *aggregated* network, ie the network found by aggregating all the nodes and their links at all times. However, this approach discards all the temporal information, and may lead to inappropriate descriptions, as very different dynamic data can give rise to the identical static graphs (Berger-Wolf and Saia, 2006). To avoid this problem, the opposite approach keeps most of the temporal information by building networks for successive time slices. Then, the mesoscopic structure of each of these slices is found *independently* for each. These structures are then connected in various ways to obtain a temporal description (Berger-Wolf and Saia, 2006 ; G. Palla et al, 2007, Rosvall and Bergstrom, 2010, Chavalarias and Cointet, 2013). By using an optimal structural description at each time slice, this method avoids the inertia of the aggregated approach. Its main drawback lies in the inherent fuzziness of the community structures, which translates into “noise” in the community detection, leading to artificial mesoscopic evolutions, with no counterpart in the real evolutions of the data. For example, rather different partitions have a very close modularity (Good et al, 2010), and minor changes in the network may lead to quite different partitions in successive time slices, which would be inadequately interpreted as major structural changes.

Several methods have been proposed to overcome the problems of these two extreme approaches (Gauvin et al. 2015, Peel and Clauset, 2014, Mucha et al, 2010, Kawadia and Sreenivasan, 2012). Here, we present “continuous partitioning”, a new approach that helps distinguishing between real trends and noise in the data. Our basic idea is to guide the community detection at time t by introducing some memory of the structure at time $t-1$, thus avoiding artificial changes induced by noise. Adapting a suggestion by Kawadia and Sreenivasan (2012), we define the distance between two successive partitions as the proportion of edges that were internal to a community at $t-1$ and connect nodes belonging to different communities at t or the reverse. We call “estranged” edges the former, and we add the latter, ie “fused” edges which used to connect distinct communities at $t-1$ but are inside a single community at t . In practice, we optimize a “continuous modularity” function Q_c , given by $Q_c = Q - (e + f)$, where Q is the usual modularity, and e and f are the proportion of estranged and fused edges. In this way, we obtain a

mesoscopic description that is able to follow the changes in real time, adapting to its instantaneous configuration, but that does not introduce the noise induced by *independently* optimizing the structure at each time step. We show the relevance of our method on the analysis of a scientific network showing the birth of a new subfield, wavelet analysis.

Dataset description To retrieve a relevant set of publications, we identified 83 key actors of early developments of wavelets. The list was established using expert advice (one of the authors, PF) and bibliographic searches. We then retrieved all their publications (from 1970 to 2012), obtaining 6,500 records from Web of Science.

Emergence and evolution of a new scientific field Studying the emergence of wavelets analysis is interesting for several reasons. First, wavelets are of particular relevance for signal/image processing, which benefited from new bridges between mathematics, physics and electrical engineering. On a more methodological ground, wavelets history is interesting because it is a recently born subfield (seminal paper in 1984), for which most actors are still around for feedback. Moreover, there exists a large corpus of publications which allows for a meaningful use of our quantitative tools. We have grouped the articles in 4-years wide time slices, separated by one year. Therefore, there is a 3-year overlap between subsequent time slices. For each slice, we first define a network by linking the articles that share 2 or more references using bibliographic coupling. We then compute the successive communities by the continuous partitioning discussed above, and by the independent maximization of modularity for each slice for comparison. Our main results are :

1. Methodological strengths of our method The figure shows that our method respects the continuity of successive partitions at almost no cost in modularity. This suggests that it succeeds in distinguishing the noise from the real network evolutions, which are more gradual. The figure also displays the time evolution of the distance between the successive networks, as measured by the sum of estranged and fused edges. There is a clear peak in year 1991, pointing to a major transition, a major change in network structure, which prevents a partition closer to the 1990 one. Note that in 1991, almost 20% of the edges are either estranged or fused from the 1990 partition. This correlates with a modularity of 0.42, which points to a weak structure of the network.

2. Successive periods in wavelets history The convergences/divergences of the different subfields can be quantified by computing the modularity of the network obtained at each time slice. Roughly speaking, a high modularity value corresponds to isolated clusters, while low values point to highly interconnected networks. The analysis shows that there are three main stages. In an initial phase (before ~ 1985), researchers work in different, relatively unrelated fields and modularity is high. Then, around 1991, wavelets appear as a common topic whose use gains momentum, defining a new, specific field that interlinks scholars, leading to a minimum in modularity. After this, modularity increases again, pointing to a new, softer divergence, as the

initial levels are not reached. Wavelets become a mature tool, that are less an object of interest *per se*, serving instead a more ancillary role within specialized communities and paving the way for new avenues of research, by developing new tools (as exemplified by the emergence of “compressed sensing”) or applying wavelets to specific, relatively unrelated, domains.

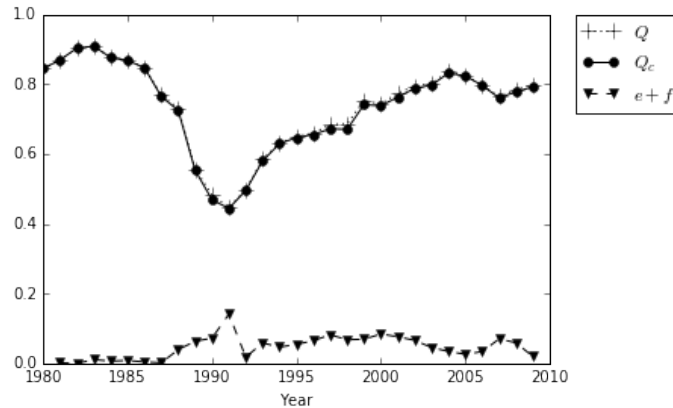


Figure Q_Qc: (upper curves) Evolution of the modularities of the partitions obtained by maximizing modularity (+, Q) or continuous modularity (O, Qc). (lower curve) Evolution of the distance between the successive networks, as measured by sum of the proportions of estranged and fused edges.

References :

- Berger-Wolf TY and J. Saia, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006, pp. 523–528
- Chavalarias D, Cointet J-P (2013) PLoS ONE 8(2): e54847. doi:10.1371/ journal.pone.0054847
- Good BH et al (2010) Physical Review E 81, 046106
- Holme, P., Saramäki, J. (2012) Physics reports, 519(3): 97-125;
- Holme P, The European Physical Journal B, vol. 88, no. 9, pp. 1–30, 2015.
- Fortunato S. Physics Reports 486, 75–174 (2010)
- Gauvin L, A. Panisson, A. Barrat, C. Cattuto, arXiv:1501.02758 (2015)
- Kawadia V, Sreenivasan S. Sci Rep 00794 (2012)
- Mucha PJ et al. Science 328, 876 (2010);
- Palla G, A.-L. Barabasi, and T. Vicsek, Nature, vol. 446, no. 7136, pp. 664–667, 2007
- Peel L and Clauset A, Proc. of the 29th International Conference on Artificial Intelligence (AAAI), 2914-2920 (2015).
- Rosvall M, Bergstrom CT (2010) PLoS ONE 5(1): e8694. doi:10.1371/journal.pone.0008694