Accepted Manuscript

Community detection in networks using self-avoiding random walks

Guilherme de Guzzi Bagnato, José Ricardo Furlan Ronqui, Gonzalo Travieso

PII: S0378-4371(18)30424-2
DOI: https://doi.org/10.1016/j.physa.2018.04.006
Reference: PHYSA 19438

To appear in: Physica A

Received date: 10 May 2017
Revised date: 21 January 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Community detection in networks using self-avoiding random walks

Guilherme de Guzzi Bagnato,* José Ricardo Furlan Ronqui, and Gonzalo Travieso†
Instituto de Física de São Carlos, Universidade de São Paulo, CP 369, 13560-970, São Carlos, SP, Brazil
(Dated: January 21, 2018)

Different kinds of random walks have proven to be useful in the study of structural properties of complex networks. Among them, the restricted dynamics of self-avoiding random walks (SAW), which visit only at most once each vertex in the same walk, has been successfully used in network exploration. The detection of communities of strongly connected vertices in networks remains an open problem, despite its importance, due to the high computational complexity of the associated optimization problem and the lack of a unique formal definition of communities. In this work, we propose a SAW-based method to extract the community distribution of a network and show that it achieves high modularity scores, specially for real-world networks. We combine SAW with principal component analysis to define the dissimilarity measure to be used for agglomerative hierarchical clustering. To evaluate the performance of this method we compare it with four popular methods for community detection: Girvan-Newman, Fastgreedy, Walktrap and Infomap using two types of synthetic networks and six well-known real-world cases.

I. INTRODUCTION

In the past few years, many dynamical processes such as percolation [1], synchronization [2] and epidemic spreading [3] have been studied in a wide variety of complex networks. Among these processes, random walks have proven to be a flexible tool to characterize and explore networks. In a random walk, a walker visits one vertex per step chosen randomly among all neighbors of the current vertex. Although simple, this process has the advantage of using only local information of the network, making it convenient when few properties of the system as a whole are known. Several properties of random walks on complex networks have been studied as, for example, scaling behavior in small world networks [4, 5], first passage time [6–8], characteristics of this dynamic in directed networks [9–11] as well as the effect of finite memory [12, 13].

Many alternative types of random walks have been proposed to optimize topological analyses [14–19]. Among these approaches, the self-avoiding walk (SAW) was shown to be more efficient in the exploration and navigation of different network structures than the traditional walker. In the SAW, the walkers cannot return to an already visited vertex, forcing them to find a new viable path through unvisited vertices. If no new vertices are available, the walk ends. Because the walker retains a memory of the path traveled, a general analytic solution is not trivial. However, some theoretical efforts yielded interesting results for small-world networks [20], Erdős-Rényi [21] and scale-free [22, 23] topologies.

One feature of complex networks that has been subject of research in several fields such as physics, biology and economy is the presence of a community structure: groups of vertices densely connected to each other and (comparatively) sparsely connected with the rest of the network [24]. In recent years, a variety of methods for identifying these groups were developed using dynamical process based on structural properties, including random walks [25–29]. In order to find the best community distribution, such algorithms as spectral methods [30, 31] and extremal optimization [32] frequently work by optimizing the quality function known as modularity [33, 34], which compares the density of edges within the communities with the expected number if the vertices were attached at random, since a random graph is not expected to have a community structure.

Despite the interdisciplinarity and the wide practical importance of finding communities for the study of e.g. metabolic process, marketing strategies and improving the routing in World Wide Web, community detection remains an open problem due to the high computational complexity of the optimization process required to uncover this structure in a network. However, many heuristic methods of modularity optimization show good agreement with peculiar relevant characteristics of real systems. The aforementioned suggests that it could be advantageous to use the high effectiveness of SAWs in exploring the network structure for the implementation of a community detection algorithm.

We use two properties of SAW for each pair of vertices $i$ and $j$: the probability for a walker departing from vertex $i$ to reach vertex $j$ before stopping and the average number of steps taken by walks that reach $j$. As is shown below, the ratio of these properties allows us to identify some structural patterns related to the community structure. The algorithm proposed in this paper is based on this information and agglomerative hierarchical clustering [35, 36]. We show that the algorithm matches or supplant the performance (with respect to modularity optimization) of the other traditional hierarchical methods used for comparison and, consequently, enhances the precision of community detection.

The paper is organized as follows. In Sec. II, we define some measurements related to SAW’s dynamic and

---

* guilherme.bagnato@usp.br
† gonzalo@ifsc.usp.br
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات