Community Detection in Complex Networks with Quantum Random Walks

Dimitris I. Tsomokos
Department of Mathematics, Royal Holloway, University of London, Egham, TW20 0EX, United Kingdom
(Dated: December 14, 2010)

Complex networks are structurally disordered systems that often display clustering behavior. The emergent clusters, also known as communities, consist of nodes that are more connected among themselves than they are connected with the rest of the network. Analyzing community structure is an important problem in network theory, with numerous applications in different fields. In this work I investigate the evolution of a continuous-time quantum random walk on a social network with benchmark community structure and show that it can be used to perform community detection.

PACS numbers: 03.67.Ac, 75.10.Jm, 89.75.Kd

Introduction — Networks are ubiquitous in both nature and society. They are routinely used to simulate a wealth of phenomena in the physical and biological sciences, as well as in sociology, finance, information and communication technologies [1,2]. In the vast majority of such applications the employed networks are inherently complex, by which we mean that there are strong fluctuations in their structural characteristics. This structural disorder is, in fact, a new type of disorder that can lead to cooperative behavior which goes beyond the one encountered in traditional condensed matter physics [3].

Historically, studies of networks have their origin in graph theory. A graph \( G \) consists of a set of vertices \( V(G) \) and a set of edges \( E(G) \), where an edge is an unordered pair of distinct vertices. Many important properties of graphs can be derived algebraically using matrices that are associated with them. For instance, the adjacency matrix \( A(G) \) of a graph \( G \) (‘network’), defined explicitly below, gives the connections between vertices (‘nodes’ or ‘sites’) according to the edge set and has been used extensively in the past [4].

In various practical applications of graph theory, such as parallel or distributed computing in computer science, it is often necessary to partition a graph into smaller subgraphs. As a result, a large body of work has been devoted to the problem of graph partitioning [5]. A similar problem in the study of networks is the identification of smaller parts, which are structurally different from the whole in the sense that they are more connected internally (between nodes on the inside) than they are connected externally (with nodes on the outside). In other words, the problem is to analyze a complex network in terms of its so-called community structure [6]. Although similar to graph partitioning, community detection is quite different in that one does not know a priori the number of communities that the network can be partitioned into. Complex networks and community detection have been approached with statistical mechanics and related toolboxes in recent years [7].

Arguably, one of the most influential algorithms for community detection is the Girvan-Newman algorithm [8]. At the heart of the algorithm lies the identification of intercommunity edges using a centrality measure known as edge betweenness. According to this measure, edges are ranked based on the number of shortest paths that pass through them. The algorithm proceeds by repeatedly removing the highest ranking edge (i.e., the one with the highest ‘traffic’) and calculating the betweenness of remaining edges until all the network communities have been isolated from one another. There are various complex networks that can serve as benchmarks, against which community detection algorithms can be tested [7]. One such standard benchmark, depicted in Fig. 1, is a social network known as Zachary’s karate club (KC) [3,10].

In this work, the use of quantum systems as dynamical probes of network structure is proposed and analyzed in a proof-of-principle implementation. In particular, I examine the evolution of a continuous-time quantum random walk [11] on a complex network with benchmark community structure, that is, the KC network of Fig. 1. In previous studies, classical random walks have been used to find distance measures and alternatives to modularity optimization [12]. Here, their quantum counterparts are employed in a dynamical setting, in which the probability of finding the particle on a node \( j \) is the only quantity needed to reveal the centrality of the node and the community in which it belongs. By defining a node affinity function, the community membership of each node can be quantified in relation to all other nodes in the network.

FIG. 1: (Color online) Community structure in the karate club (KC) network [9] with \( N = 34 \) nodes. The two main communities, centered around nodes 1 and 34, are indicated by squares and circles, respectively. Colors correspond to the best partition found by optimizing the Girvan-Newman algorithm. Reprinted from Ref. [10] (©2004, IOP Publishing and SISSA).
Model and Definitions — A complex network \( G \), composed of \( N \) vertices and \( K \) edges, can be described by its adjacency matrix \( A(G) \). For an undirected network without loops this matrix is given by

\[
A_{ij} = w_{ij} \text{ if } (i, j) \in E(G), \\
A_{ij} = 0 \text{ otherwise.}
\]

If the network is unweighted we let \( w_{ij} = 1 \), and this is indeed the case considered here. At \( t = 0 \) the initial state of the network is encoded in an \( N \times 1 \) vector, \( |\Psi(0)\rangle \). At later times the evolved state is

\[
|\Psi(t)\rangle = \exp (-iAt) |\Psi(0)\rangle. \tag{1}
\]

The probability of finding the quantum random walker on a node \( j \) at time \( t \) is

\[
P_j(t) = |\Psi_j(t)|^2 \tag{2}
\]

where \( \Psi_j \) denotes the \( j \)-th element of the state vector \( |\Psi(t)\rangle \) and we have \( \sum_j P_j(t) = 1 \). The time-averaged quantity

\[
\bar{P}_j = \frac{1}{T} \int_0^T P_j(t) dt \tag{3}
\]

gives the mean probability of finding the walker on node \( j \). In what follows, \( P_j \) plays a crucial role and it is referred to as the population of node \( j \).

The calculation of these quantities proceeds from the full spectrum of the adjacency matrix \( A \) of \( G \). If the spectrum of \( A \) is \( \{\epsilon_n, |V_n\rangle\} \) for \( n = \{1, 2, \ldots, N\} \), then the evolution operator of Eq. (1), \( \exp (-iAt) \), can be expressed as \( \sum_n \exp(-i\epsilon_n t)|V_n\rangle \langle V_n| \), where \( |V_n\rangle \) is the conjugate eigenvector. Given some initial state \( |\Psi(0)\rangle \neq |V_n\rangle \), which obeys the condition \( \langle \Psi(0)|\Psi(0)\rangle = \sum_j |\Psi_j(0)|^2 = 1 \), the evolved state is obtained using this decomposition of \( \exp (-iAt) \).

Interestingly, it was recently shown that the eigenvalue spectrum of \( A \) gives a clear indication of the number of communities present in a network \([13]\). In particular, for a network with \( N \) nodes and \( \mathcal{N} \) communities, there are typically \( \mathcal{N} \) eigenvalues that are significantly larger than the remaining \( (N - \mathcal{N}) \) eigenvalues. Indeed, the eigenvalue spectrum of the adjacency matrix for the KC network of Fig. \( 1 \) indicates the presence of two main communities \([14]\). However, as shown in the present work, when the full spectrum of \( A \) is analyzed in the quantum setting of Eq. (1), then it can also be used to reveal the actual community structure of a network.

Furthermore, it may be useful to clarify that, formally, a continuous-time quantum random walk \([11]\) makes use of the Laplacian matrix \([4]\) of a given network. Here we use the adjacency matrix instead, which generates the same quantum dynamics as the Laplacian only for regular graphs. In the case of complex (structurally disordered) networks the two evolutions are different, in general, but of course both of them can be considered as quantum random walks.

In terms of a physical model, the adjacency matrix \( A \) is in fact the effective Hamiltonian of an \( XY \)-interaction spin model restricted to the single-excitation subspace. This model has been studied extensively in the field of quantum communication with spin chains and lattices \([15] [16]\). Therefore, in this context, the walker is a quantum excitation (i.e., a quasiparticle) diffusing in the network according to Eq. (1) and so, in the following, the term excitation is sometimes used to describe the quantum random walker.

Populations Concentrate in Central Nodes — One of the most basic structural characteristics of a complex network is the centrality of its nodes. A straightforward measure is the degree centrality,

\[
C_j = \frac{d(j)}{N - 1},
\]

where \( d(j) \) is the degree of a node \( j \). For \( C(j) = 0 \) the node is isolated; and for \( C(j) = 1 \) the node is connected with every other node in the network.

In order to probe the centrality of nodes in the KC network using our formalism, we begin by assuming that there is equal \( a \) priori probability \( p_0 \) of measuring the excitation in any given node. Therefore the initial state is

\[
|\Psi(0)\rangle = |p_0 \cdots p_0\rangle,
\]

where \( p_0 = 1/\sqrt{N} \). The evolved state \( |\Psi(t)\rangle \) is obtained numerically for a period of time \( t \in [0, T] \), in time-steps \( \delta t \ll T \), and consequently the populations are calculated as prescribed by Eq. (3).

By comparing the population of each node with its degree centrality, one sees that populations tend to flow into highly-connected nodes. The result is presented in Fig. 2 where both the degree centralities \( C_j \) and the populations \( P_j \) are shown. It is clear that the population distribution is correlated with the degree centrality of the nodes.

![FIG. 2: (Color online) Degree centrality \( C_j \) (line of squares) and population \( P_j \) (line of circles) for each node \( j = \{1, 2, \ldots, 34\} \) in the KC network of Fig. 1. Parameters used for the numerical simulation: \( T = 100\pi \) (dimensionless units) and \( \delta t = 10^{-3}T \).](image-url)
Flow of Populations after Edge Removal — We are now in a position to pose the central question of this work, namely, ‘If an edge is removed, how do populations move?’

We observed that populations tend to flow out of peripheral nodes and into central nodes. So it is reasonable to expect that population hubs are formed around central nodes and that these hubs are sustained by intracommunity edges, which connect members of a single community. As a result, the removal of an edge strictly from the interior of a community weakens the hub and therefore the population of the corresponding community as a whole decreases (while of course the population of the rest of the network increases). The working hypothesis is that the intercommunity edges, which connect two different communities, should be identifiable by their deviation from this generically expected behavior.

The hypothesis has been tested and verified numerically in the KC network by removing edges and recalculating populations after every removal operation. Typically, if the edge belongs to a community, populations flow out of that particular community and into neighboring ones. But if the edge removed connects two communities, the populations of both hubs inside these communities are increased (while of course the populations of the two nodes, which were previously connected, are reduced). Deviations from these two general cases can occur for certain edges belonging to a sub-community, such as edges connecting the yellow squares in Fig. 1. However these instances can be identified by the unconventional population flows that they give rise to, as explained shortly.

Three illustrative examples are presented in Fig. 3. In Fig. 3(a) (b) the removed edge belongs to the community indicated by squares (circles) in the KC network of Fig. 1 after its removal the populations of the squares (circles) are reduced, as seen by the green (red) diamonds. By contrast, in Fig. 3(c) the edge lies in-between the communities of squares and circles; and after its removal the population hubs in both communities are increased.

Therefore, edges lying in-between communities are identified by the fact that their removal increases the population hubs of both connected communities. This is in distinct contrast with the removal of an intracommunity edge, which typically increases populations strictly outside its own community. However, a more complicated flow may occur when we remove an edge from a sub-community, or an edge that connects a sub-community with the rest of the community. Examples of such edges are (5,11) and (1,11). In some of these special cases, populations in the sub-community increase after the edge removal, while populations in the rest of the community decrease. Such unconventional population flows can indeed be used to flag the presence of a sub-community.

Node Affinity and Fuzzy Communities — It is possible to have overlapping, or fuzzy, communities when a node is associated with two or more clusters (see, e.g., Refs. [7, 17]). It would therefore be useful to have a measure which quantifies the community membership of a node. To arrive at such a measure we begin by asking, ‘What is the probability that two nodes \(i\) and \(j\) belong to the same community?’

\[
\alpha_{ij} = \frac{1}{K} \sum_{k=1}^{K} \theta_j(k) \theta_j(k). \tag{4}
\]

The results of Fig. 3 point to the fact that community structure is revealed through relative population flows, after each edge removal operation. The crucial quantity is the direction of population flow (i.e., whether it is flowing into or out of a node) and not the absolute value of the flow. So we let the direction of flow on node \(j\), after an edge \(k \in [1, K]\) has been removed, be given by \(\theta_j(k) = \pm 1\), where \(+1\) \((-1\) denotes population flowing into (out of) the node. We define

\[
\alpha_{ij} = \frac{1}{K} \sum_{k=1}^{K} \theta_j(k) \theta_j(k). \tag{4}
\]

which takes values between \(-1\) and \(1\). Whenever \(\alpha_{ij} > 0\)
The node affinity function for the KC network is presented in Fig. 4. Interestingly, nodes with high affinity with nodes 2 to 8 and 11 to 14 (among others), while it has low affinity with nodes $\{9, 10, 15, 16, 19, 21\}$ and 23 to 34. Interestingly, nodes 3 and 20 have the lowest affinity with other nodes, which implies they have a fuzzy community membership (this verifies the recent results of Liu [17] for the KC network, obtained with an entirely different methodology).

Conclusions and Outlook — The results presented here demonstrate that community structure and other structural characteristics of a complex network can be revealed through the dynamics of a continuous-time quantum random walk. In particular, we have examined the case where, initially, there is equal a priori probability to find the quantum random walker (excitation) on any node in the KC network of Fig. 4 which has benchmark community structure and is used to demonstrate the processes involved. The system is let to evolve for time $T$ and the mean probability of finding the excitation on a given node is calculated. We find that this mean probability (population) is correlated with the centrality of the node. Populations tend to concentrate in well-connected nodes and, as a result, population hubs are formed around such nodes.

A clear physical picture emerges when we observe that populations are ‘fed into’ hubs via the intracommunity edges which support them. Therefore, our hypothesis has been that the removal of an edge from inside a community $A$ typically reduces the population of $A$ and increases the populations outside $A$; while the removal an edge connecting $A$ and $B$ increases the population hubs of both communities. By defining a node affinity function we have quantified the likelihood that two nodes belong to the same community and shown that the function can probe fuzzy communities as well.

In future work the scheme will be applied to other networks of increasing size and complexity, and tested against benchmark algorithms. It is envisaged that a mutually beneficial exchange of concepts and methods can take place between quantum information science and community detection.

Acknowledgments — This work was supported by EPSRC-GB under Grant No. EP/G045771/1. I would like to thank Tobias Osborne for very useful conversations.

[14] The method of Ref. [13] is the following: (i) calculate the eigenvalues of $A$, $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_N$; (ii) calculate the gaps between them, $\Delta_k = \epsilon_k - \epsilon_{k+1}$, for all $k = 1, 2, \ldots, N - 1$; (iii) pick the largest eigengap $\Delta_k$. Then the value of $k$ gives the number of communities, i.e., $N = c$ if and only if $\Delta_k = \max(\Delta_k)$. For the KC network [9] we have $\max(\Delta_k) = \Delta_2$.