Change Point Detection in a Dynamic Stochastic Blockmodel

Peter Wills and François G. Meyer(✉)

Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA
fmeyer@colorado.edu

Abstract. We study a change point detection scenario for a dynamic community graph model, which is formed by adding new vertices and randomly attaching them to the existing nodes. The goal of this work is to design a test statistic to detect the merging of communities without solving the problem of identifying the communities. We propose a test that can ascertain when the connectivity between the balanced communities is changing. In addition to the theoretical analysis of the test statistic, we perform Monte Carlo simulations of the dynamic stochastic blockmodel to demonstrate that our test can detect changes in graph topology, and we study a dynamic social-contact graph.

Keywords: Change detection · Dynamic stochastic blockmodel · Graph distance

1 Introduction

Some of the most well-known empirical network datasets reflect social connective structure between individuals, often in online social network platforms such as Facebook and Twitter. These networks exhibit structural features such as communities and highly connected vertices, and can undergo significant structural changes as they evolve in time. Examples of such structural changes include the merging of communities, or the emergence of a single user as a connective hub between disparate regions of the graph.

The main contribution of this work is a rigorous analysis of a dynamic community graph model, which we call the dynamic stochastic blockmodel. Models of dynamic community networks have recently been proposed. The simplest incarnation of such models, the dynamic stochastic blockmodel, is the subject of our study.

This model is formed by adding new vertices, and randomly attaching them to the existing nodes. We circumvent the problem of decomposing each graph into communities, and propose instead a test that can ascertain when the connectivity between the balanced communities is changing. Because the evolution of the graph is stochastic, one expects random fluctuations of the graph topology.

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2019, SCI 881, pp. 211–222, 2020.
https://doi.org/10.1007/978-3-030-36687-2_18

We propose an hypothesis test to detect the abnormal growth of the balanced stochastic blockmodel.

The stochastic blockmodel represents the quintessential exemplar of a network with community structure. In fact, it is shown in [28] that any sufficiently large graph behaves approximately like a stochastic blockmodel. This model is also amenable to a rigorous mathematical analysis, and is indeed at the cutting edge of rigorous probabilistic analysis of random graphs [1].

2 Graph Models

We recall the definition of the two-community stochastic blockmodel [1].

Definition 1. *Let $n \in \mathbb{N}$, and let $p, q \in [0,1]$. We denote by $\mathsf{SBM}(n,p,q)$ the probability space formed by the graphs defined on the set of vertices $[n]$, constructed as follows. We split the vertices $[n]$ into two communities C_1 and C_2, formed by the odd and the even integers in $[n]$ respectively. We denote by $n_1 = \lfloor (n+1)/2 \rfloor$ and $n_2 = \lfloor n/2 \rfloor$ the size of C_1 and C_2 respectively. Edges within each community are drawn randomly from independent Bernoulli random variables with probability p. Edges between communities are drawn randomly from independent Bernoulli random variables with probability q.*

2.1 The Dynamic Stochastic Blockmodel

Several dynamic stochastic blockmodels have been proposed in recent years (e.g., [10,11,18,22,27,29,31], and references therein). Existing dynamic stochastic block models assume that the number of nodes is fixed, and that community membership is random. Some authors propose a Markovian model for the community membership [30,31], while others assume the sequence of graphs are independent realizations in time [5]. Our work is more similar to that of [4], where the authors study changes in the dynamics of a preferential attachment model, the size of which grows as a function of time.

Similarly, we investigate a growing model of a stochastic block model, and we are interested in the regime of large graphs ($n \to \infty$), where the probabilities of connection, within each community, p_n, and across communities, q_n, go to zero as the size of the graph, n, goes to infinity.

In order to guarantee that at each time n we study the growth of a graph $\sim \mathsf{SBM}(n, p_n, q_n)$, we cannot simply assume that the graphs $G_1 = (V_1, V_1), \ldots, G_n = (V_n, E_n)$ form a sequence of nested subgraphs, where we would have $V_1 \subset \cdots \subset V_n$ and $E_1 \subset \cdots \subset E_n$. Instead. our study focuses on the transition between a random realization $(V_n, E_n) \sim \mathsf{SBM}(n, p_n, q_n)$ and the graph formed by adding a new node $n+1$ and random edges to (V_n, E_n).

Formally, the dynamic stochastic blockmodel is defined recursively (see Table 1). G_1 is formed by a single vertex. We assume that we have constructed $G_1, \ldots, G_n$ and we proceed with the construction of G_{n+1}. First, we replace G_n by a graph $H_n \sim \mathsf{SBM}(n, p_n, q_n)$, and we consider the graph formed by adding a new node $n+1$ (assigned to either C_1 or C_2 according to the parity of n), and we define

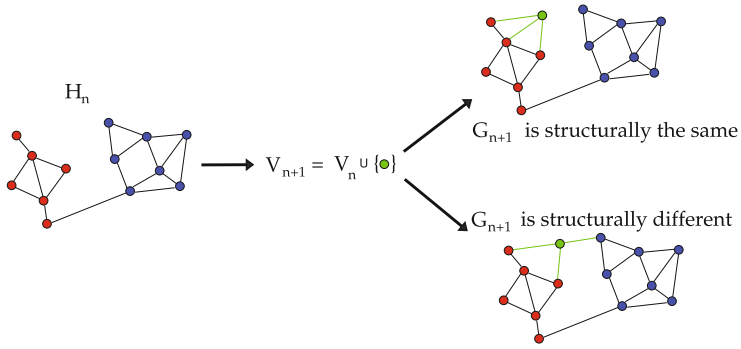

Fig. 1. Left: the stochastic blockmodel $H_n = (V_n, E_n)$ is comprised of two communities C_1 (red) and C_2 (blue). A new vertex (green) is added to V_n, and random edges are created between $n+1$ and vertices in V_n. This leads to a new set of edges, E_{n+1}, and the corresponding new graph G_{n+1} defined by (2)

$$V_{n+1} \triangleq V_n \cup \{n+1\}\,. \tag{1}$$

Random edges are then assigned from $n+1$ to each vertex in the same community with probability p_n and to each vertex of the opposite community with probability q_n. This leads to a new set of edges, E_{n+1}, and the corresponding graph (see Fig. 1),

$$G_{n+1} \triangleq (V_{n+1}, E_{n+1})\,. \tag{2}$$

We note that H_n is different from G_n; indeed, G_n was created by adding a node and some edges to the graph $H_{n-1} \sim \mathsf{SBM}(p_{n-1}, q_{n-1})$, whereas H_n is a realization of $\mathsf{SBM}(p_n, q_n)$. Table 1 summarizes the construction of the sequence $G_1, G_2, \ldots$

Table 1. Row n depicts the construction of G_{n+1} as a function of the random seed graph H_n. The distance (last column) is always defined with respect to the seed graph H_n on the vertices $1, \ldots, n$ that led to the construction of G_{n+1}.

Time index n	Probabilities of connection	H_n = seed graph at time n	Growth sequence to generate G_{n+1}	Definition of the graph distance
0		$\emptyset$	$\emptyset \to \boxed{G_1}$	0
1	$\{p_1, q_1\}$	$H_1 \sim \mathsf{SBM}(1, p_1, q_1)$	$H_1 \to \boxed{G_2}$	$\mathrm{d_{rp}}(H_1, G_2)$
2	$\{p_2, q_2\}$	$H_2 \sim \mathsf{SBM}(2, p_2, q_2)$	$H_2 \to \boxed{G_3}$	$\mathrm{d_{rp}}(H_2, G_3)$
⋮	⋮	⋮	⋮	⋮

We conclude this section with the definition of the expected degrees and the number of across-community edges, k_n.

Definition 2 (Degrees and number of across-community edges). *Let $G \sim \mathsf{SBM}(n, p, q)$. We denote by $\overline{d_{n1}} = pn_1$ the expected degree within community C_1, and by $\overline{d_{n2}} = pn_2$ the expected degree within community C_2. We denote by k_n the binomial random variables that counts the number of cross-community edges between C_1 and C_2.*

Because asymptotically, $n_1 \sim n_2$, we ignore the dependency of the expected degree on the specific community when computing asymptotic behaviors for large n. More precisely, we loosely write $1/\overline{d_n}$ when either $1/\overline{d_{n1}}$ or $1/\overline{d_{n2}}$ could be used.

3 The Resistance Perturbation Distance

In order to study the dynamic evolution of the graph sequence, we focus on changes between two successive time steps n and $n+1$. These changes are formulated in terms of changes in connectivity between G_{n+1} and the seed graph H_n (see Table 1).

To construct the statistic that can detect the merging of communities without identifying the communities, we use the resistance perturbation distance [17]. This graph distance can be tuned to quantify configurational changes that occur on a graph at different scales: from the local scale formed by the local neighbors of each vertex, to the largest scale that quantifies the connections between clusters, or communities [17] (see [2,6] for recent surveys on graph distances, and [19] for a distance similar to the resistance perturbation distance).

The Effective Resistance. For the sake of completeness, we review the concept of effective resistance (e.g., [7–9,12]). Given a graph $G = (V, E)$, we transform G into a resistor network by replacing each edge e by a resistor with conductance w_e (i.e., with resistance $1/w_e$). The **effective resistance** between two vertices u and v in V is defined as the voltage applied between u and v that is required to maintain a unit current through the terminals formed by u and v.

To simplify the discussion, we will only consider graphs that are connected with high probability. All the results can be extended to disconnected graphs as explained in [17].

Definition 3 (The resistance perturbation distance). *Let $G^{(1)} = (V, E^{(1)})$ and $G^{(2)} = (V, E^{(2)})$ be two graphs defined on the same vertex sets V. Let $R^{(1)}$ and $R^{(2)}$ denote the effective resistances of $G^{(1)}$ and $G^{(2)}$ respectively. We define the resistance-perturbation distance to be*

$$\mathrm{d_{rp}}\left(G^{(1)}, G^{(2)}\right) \triangleq \sum_{u \in V} \sum_{v \in V, v \neq u} \left| R^{(1)}_{u\,v} - R^{(2)}_{u\,v} \right|. \tag{3}$$

The resistance-perturbation distance cannot be used to compare graphs defined on different vertex sets, $V^{(1)}$ and $V^{(2)}$. If $V^{(1)}$ and $V^{(2)}$ share many nodes, then we can compute the restriction of the perturbation distance on the intersection $V^{(1)} \cap V^{(2)}$. In the following we compare two graphs H_n and G_{n+1} that share all the nodes but one newly added node. We therefore extend the definition of the perturbation distance as follows.

Definition 4 (Extension of the resistance perturbation distance). *Let $H_n = (V_n, E_n) \sim \mathsf{SBM}(n, p_n, q_n)$, and let $G_{n+1} \triangleq (V_{n+1}, E_{n+1})$, defined by (2). We define the resistance-perturbation distance between H_n and G_{n+1} as follows*

$$\mathrm{d_{rp}}\left(H_n, G_{n+1}\right) \triangleq \sum_{u \in V_n} \sum_{v \in V_n, v \neq u} \left| R^{(1)}_{u\,v} - R^{(2)}_{u\,v} \right|. \tag{4}$$

Because $n+1$ did not exist at time n, it is not meaningful to compute $R_{v(n+1)}$, for $v \in V_n$ Therefore we only compute the effective resistances for $u, v \in V_n$ in (4). In the remainder of the paper, we use the notation $\mathrm{d_{rp}}$ to denote the extended resistance perturbation distance defined by (4).

4 Main Result

Figure 1 illustrates the statement of the problem. As a new vertex (shown in green) is added to the graph H_n the connectivity between the communities can increase, if edges are added between C_1 and C_2, or the communities can remain separated if no across-community edges are created. If the addition of the new vertex promotes the merging of C_1 and C_2, then we consider the new graph G_{n+1} to be *structurally different* from G_n, otherwise G_{n+1} remains *structurally the same* as G_n (see Fig. 1).

As explained in Theorem 1, the resistance perturbation distance, between time n and $n+1$, (see Table 1) measured by $\mathrm{d_{rp}}(H_n, G_{n+1})$ (defined by (4)) is able to distinguish between connectivity changes within a community and changes across communities.

Theorem 1 (The statistic under the null and alternative hypotheses). *Let $H_n = (V_n, E_n) \sim \mathsf{SBM}(n, p_n, q_n)$ with $p_n = \omega\left(\log n/n\right)$, and $p_n/n < q_n < (p_n/n)^{3/4}$. Let G_{n+1} be the graph generated according to the dynamic stochastic blockmodel described by (2) and Table 1.*

To test the hypothesis $k_{n+1} = k_n$ (null hypothesis) versus $k_{n+1} > k_n$ (alternative hypothesis) we use the statistic Z_n defined by

$$Z_n \triangleq \frac{p_n}{4} \mathbb{E}\left[\mathrm{d_{rp}}\left(H_n, G_{n+1}\right)\right] - 1. \tag{5}$$

The expected value of the statistic $\mathbb{E}[Z_n]$ *is given by*

$$\mathbb{E}[Z_n] = \begin{cases} \mathcal{O}\left(\frac{1}{\sqrt{d_n}}\right), & \text{conditioned on } k_{n+1} = k_n (null) \\ \frac{2p_n}{n^2 q_n^2} + \mathcal{O}\left(\frac{1}{\sqrt{d_n}}\right), & \text{conditioned on } k_{n+1} > k_n (alternative). \end{cases} \tag{6}$$

The theoretical analysis of the dynamic stochastic block model $\mathsf{SBM}(n, p_n, q_n)$, provided by Theorem 1, reveals that if one could ignore the within-community random connectivity changes, which have size $\mathcal{O}\left(1/\sqrt{d_n}\right)$, then one should always be able to detect the addition of across-community edges using the global metric provided by the test statistic Z_n. The condition $q_n < (p_n/n)^{3/4}$ therefore guarantees that within-community connectivity changes do not obfuscate across-community connectivity changes triggered by the increase in across-community edges.

Without loss of generality, we consider that the new node $n+1$ is added to C_2 (C_1 and C_2 play symmetric roles). The main result relies on the following two ingredients.

1. the community C_2 is approximately an Erdős-Rényi graph $\mathsf{SBM}(n_2, p_n)$, wherein the effective resistance R_{uv} concentrates around $2/(n_2 p_n)$ [23];
2. the effective resistance between $u \in C_1$ and $v \in C_2$ depends mostly on the bottleneck formed by the k_n across-community edges, $R_{uv} \approx 1/k_n$ [14,15], and the number of across-community edges, k_n, concentrates around $q_n n_1 n_2$.

Under the null hypothesis, about $n_2 p_n$ nodes in C_2 will become incident to the new edges created by the addition of node $n+1$. For each of these nodes, the new degree d'_u becomes $d_u + 1$ w.h.p., and therefore $R'_{uv} = R_{uv} - 1/d_u^2$, for all $v \in C_2$. By symmetry, $R'_{uv} = R_{uv} - 1/d_v^2$, for all $u \in C_2$, and the total perturbation for $u \in C_2, v \in C_2$ is $\approx 2n_2 n_2 p_n / \overline{d_n}^2 = 2/p_n$. We derive the same estimate for the perturbation $R'_{uv} - R_{uv}$ for $u \in C_1, v \in C_2$ or $u \in C_2, v \in C_1$. We conclude that $\mathrm{d_{rp}} \approx 4/p_n$ under the null hypothesis.

Under the alternative hypothesis $k_{n+1} = k_n + 1$ w.h.p., and thus $\Delta R_{uv} \approx -1/k_n^2 \approx -1/(n_1 n_2 q_n)^2$. This perturbation affects every pair of node (u, v) where $u \in C_1$ and $v \in C_2$, therefore $\mathrm{d_{rp}} \approx 2/(n_1 n_2 q_n^2) = 8/(n q_n)^2$. There is an additional term of order $\mathcal{O}\left(1/p_n^2\right)$ that accounts for the changes in effective resistance within C_2 (the community wherein node $n+1$ is added).

In order to estimate the noise term, $\mathcal{O}\left(1/\sqrt{d_n}\right)$, we need to construct estimates of the effective resistance that are more precise than those that can be found elsewhere (e.g., [23], but see [21] for estimates similar to ours, obtained with different techniques). The full detailed rigorous proof of Theorem 1 is provided in the supplementary material [25]; we give in the following the key steps.

Proof (Proof of Theorem 1). The proof proceeds in two steps: we first analyze the null hypothesis, and then the alternative hypothesis. Due to space limitation

we only present the alternative case, $k_{n+1} > k_n$. We have

$$\mathbb{E}\left(\mathrm{d_{rp}}\left(G_n, G_{n+1}\right)\right) = \sum_{u\in C_1}\sum_{v\in C_1}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) + \sum_{u\in C_2}\sum_{v\in C_2}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) + \sum_{u\in C_1}\sum_{v\in C_2}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) + \sum_{u\in C_2}\sum_{v\in C_1}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) \quad (7)$$

The first sum in (7) is equal to

$$\sum_{u\in C_1}\sum_{v\in C_1}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) = \frac{2n_1}{{d_n}^2}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{d_n}}\right)\right). \quad (8)$$

Similarly, we have

$$\sum_{u\in C_2}\sum_{v\in C_2}\mathbb{E}\left(|R_{uv} - R'_{uv}|\right) = n_2(n_2-1)2p_n\left(\frac{1}{{d_n}^2} + \mathcal{O}\left(\frac{1}{{d_n}^{5/2}}\right)\right). \quad (9)$$

The estimates (8) and (9), which quantify the connectivity within both communities, are oblivious to the increase in the across-community connectivity ($k_{n+1} > k_n$). We need the third and fourth terms in (7), which are significantly affected by the increase in across-community edges, to detect a change in the effective resistance. Indeed, we have

$$\sum_{u\in C_1}\sum_{v\in C_2}\mathbb{E}\left(|R_{uv} - R'_{u_{k+1}v}|\right) = \frac{1}{p_n}\left(1 + \frac{p_n}{n_1 n_2 q_n^2} + \mathcal{O}\left(\frac{1}{\sqrt{d_n}}\right)\right). \quad (10)$$

The symmetric case where $u \in C_2$ and $v \in C_1$ leads to the same exact expression. Finally, we can assemble the expected resistance perturbation distance by combining the terms (8), (9), (10), and we obtain the advertised result.

We conclude the proof with the condition on q_n that guarantees that Z_n can detect the alternative hypothesis. As soon as $q_n < (p_n/n)^{3/4}$, the statistic Z_n under the alternative hypothesis is larger than the noise term $\mathcal{O}\left(1/\sqrt{d_n}\right)$. The theoretical condition on q_n will be confirmed experimentally in the next section.

The proofs of (8), (9), and (10) are rather technical and are provided in the supplementary material [25]. □

5 Experiments

Synthetic Experiments. Figure 2 shows numerical evidence supporting Theorem 1. The experiment involves Monte Carlo simulations of the dynamic stochastic blockmodel for 64 random realizations for each q_n. The empirical distribution of Z_n is computed under the null hypothesis (green line) and the alternative hypothesis (red line). The theoretical estimate given by (7) under the alternative hypothesis is also displayed (blue line). The size of the graph is $n = 2,048$,

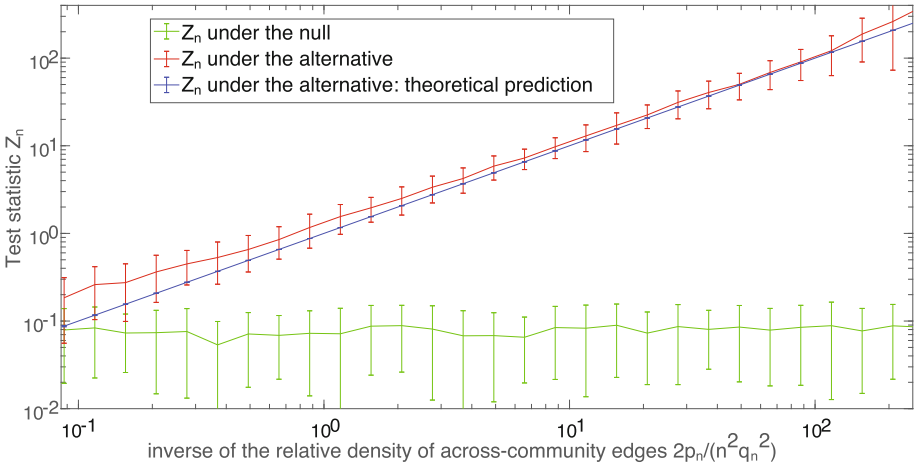

Fig. 2. Statistic Z_n defined by (5) computed under the null hypothesis (green line) and the alternative hypothesis (red line) for several values of the inverse across-community edge density, $2p_n/(n^2q_n^2)$. The theoretical estimate of Z_n under the alternative hypothesis, given by (6), is displayed as a blue line.

the density of edges is $p_n = \log^2(n)/n$, and the across-community edge density ranges from $q_{\max} = \sqrt{2}\,(p_n/n)^{3/4}$ down to $q_{\min} = q_{\max}/100$. For each value of q_n, we display the statistic Z_n as a function of the inverse across-community edge density, $2p_n/(n^2q_n^2)$. As the inverse density of across-community edges increases, the statistic Z_n can more easily detect the alternative hypothesis. The theoretical analysis, provided by (6), is confirmed: as q_n becomes larger than $\mathcal{O}\left((p_n/n)^{3/4}\right)$, the statistic Z_n computed under the null and alternative hypotheses merge. The across-community edge density q_n becomes too large for the global statistic Z_n to "sense" perturbations triggered by connectivity changes between the communities. The expected value $\mathbb{E}[Z_n] = 2p_n/(nq_n)^2$ under the alternative hypothesis becomes smaller than the noise term $\mathcal{O}\left(1/\sqrt{d_n}\right)$, and the test statistic Z_n fails to detect the alternative hypothesis.

Analysis of a Primary School Face to Face Contact. In this section we provide an experimental extension of Theorem 1, wherein there are 10 communities, but the number of nodes, N, is fixed. The across-community and within-community edge densities are rapidly fluctuating as a function of time n. Our goal is to experimentally validate the ability of the resistance-perturbation distance to detect significant structural changes between the communities, while remaining impervious to random changes within each community.

The data are part of a study [20] where RFID tags were used to record face-to-face contact between students in a primary school. Events punctuate the school day of the children, and lead to fundamental topological changes in the

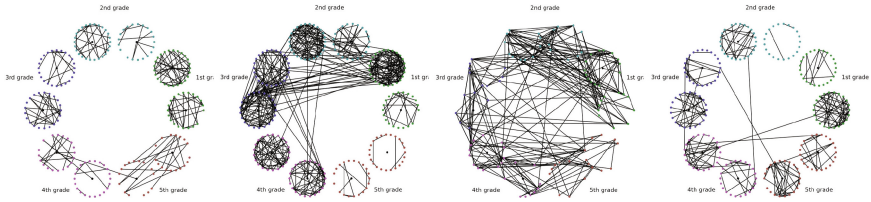

Fig. 3. Left to right: snapshots of the face-to-face contact network at 9:00 a.m., 10:20 a.m., 12:45 p.m., and 2:03 p.m.

contact network (see Fig. 3). The school is composed of ten classes: each of the five grades (1 to 5) is divided into two classes (see Fig. 3). Each class forms a community of connected students; classes are weakly connected.

During the school day, events such as lunch periods (12:00 p.m.– 1:00 p.m. and 1:00 – 2:00 p.m.) and recess (10:30 – 11:00 a.m. and 3:30 – 4:00 p.m), trigger significant increases in the number of links between the communities, and disrupt the community structure (see Fig. 3). The construction of the dynamic graphs proceeds as follows. We divide the school day into $N = 150$ time intervals of $\Delta t \approx 200$ s. We denote by $t_i = 0, \Delta t, \ldots, (N-1)\Delta t$, the corresponding temporal grid. For each t_i we construct an undirected unweighted graph G_{t_i}, where the $n = 232$ nodes correspond to the 232 students in the 10 classes, and an edge is present between two students and if they were in contact (according to the RFID tags) during the time interval $[t_{i-1}, t_i)$.

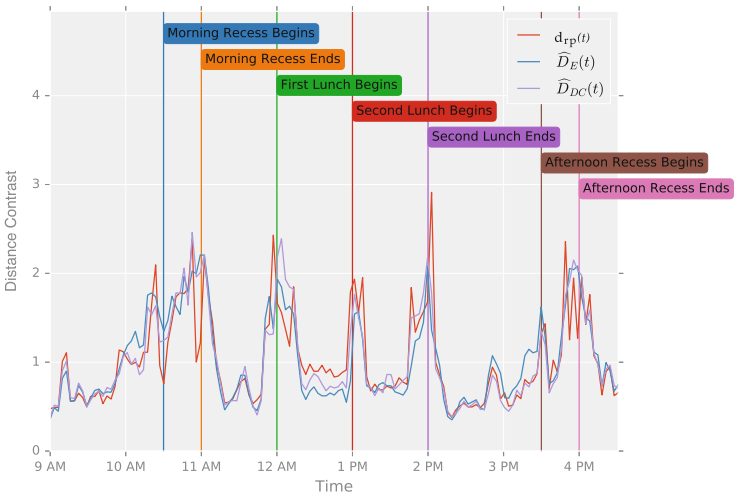

Fig. 4. Primary school data set: resistance perturbation distance d_{rp}, edit distance $\widehat{D}_E$, and DeltaCon distance $\widehat{D}_{DC}$

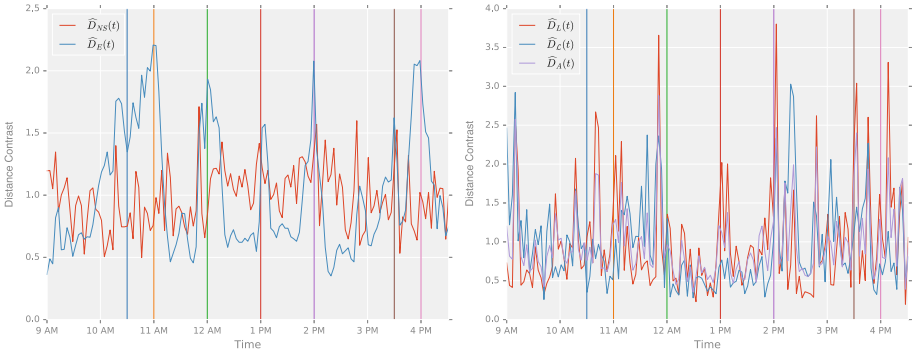

Fig. 5. Primary school data set: NetSimile $\widehat{D}_{NS}$ and edit distance $\widehat{D}_E$ (left); combinatorial Laplacian $\widehat{D}_L$, normalized Laplacian $\widehat{D}_{\mathcal{L}}$, and adjacency $\widehat{D}_A$ (right).

The purpose of the analysis is to assess whether distances can detect changes in the topology coupled with the hidden events that control the network connectivity. We are also interested to verify if distances are robust against random changes within each classroom that do not affect the communication between the classes. We compare the resistance perturbation distance $\mathrm{d_{rp}}$ to the following distances: (1) DeltaCon distance $\widehat{D}_{DC}$ [13], (2) NetSimile distance $\widehat{D}_{NS}$ [3], (3) edit distance $\widehat{D}_E$, (4) three spectral distances: combinatorial Laplacian $\widehat{D}_L$, normalized Laplacian $\widehat{D}_{\mathcal{L}}$, and adjacency $\widehat{D}_A$. The spectral distance between graph G and G' is the ℓ_2 norm of the difference between the two spectra, $\{\lambda_i\}$ and $\{\lambda'_i\}$ of the corresponding matrices [26]. For each distance measure d, we define a normalized distance contrast

$$\widehat{D}(t_i) = d(G_{t_{i-1}}, G_{t_i})/\overline{D},$$

where $\overline{D} = N^{-1} \sum_i d(G_{t_{i-1}}, G_{t_i})$. All experiments were conducted using the `NetComp` library, which can be found on GitHub at [24].

Figure 4 displays the normalized temporal differences for the resistance distance, edit distance, and DeltaCon distance. The stochastic variability in the connectivity appreciably influence the high frequency (fine scale) eigenvalues; spectral distances, which are computed using all the eigenvalues, lead to very noisy estimates of the temporal differences (see Fig. 5). NetSimile is also significantly affected by these random fluctuations.

The volume of the dynamic network changes rapidly, and the edit distance can reliably monitor these large scale changes. However, it entirely misses the significant events that disrupt the graph topology: onset and end of morning recess, onset of first lunch, end of second lunch (see Fig. 4). The resistance distance can detect subtle topological changes that are coupled to latent events that dynamically modify the networks, while remaining impervious to random local changes, which do not affect the large scale connectivity structure (see Fig. 4).

6 Discussion

We note that the condition on q_n in Theorem 1 guarantees that the communities could be recovered using other techniques (e.g., spectral clustering). Our global approach, which does not require the detection of the communities can be computed efficiently (at a cost that is comparable to fast spectral clustering algorithms). Indeed, we have developed in [17] fast (linear in the number of edges) randomized algorithms that can quickly compute an approximation to the d_{rp} distance (see [16] for the publicly available codes). In the context of streaming graphs, we described in [17] algorithms to compute fast updates of the d_{rp} distance when a small number of edges are added, or deleted.

We are currently exploring several extensions of the current model. The scenario of the primary school dataset, wherein the graph size is fixed and a latent process controls the addition and deletion of edges is an important extension of the current model.

Acknowledgements. F.G.M was supported by the National Science Foundation (CCF/CIF 1815971), and by a Jean d'Alembert Fellowship.

References

1. Abbe, E., Bandeira, A.S., Hall, G.: Exact recovery in the stochastic block model. IEEE Trans. Inf. Theor. **62**(1), 471–487 (2016)
2. Akoglu, L., Tong, H., Koutra, D.: Graph based anomaly detection and description: a survey. Data Min. Knowl. Discov. **29**(3), 626–688 (2015). https://doi.org/10.1007/s10618-014-0365-y
3. Berlingerio, M., Koutra, D., Eliassi-Rad, T., Faloutsos, C.: NetSimile: a scalable approach to size-independent network similarity. CoRR abs/1209.2684 (2012). http://dblp.uni-trier.de/db/journals/corr/corr1209.html#abs-1209-2684
4. Bhamidi, S., Jin, J., Nobel, A., et al.: Change point detection in network models: preferential attachment and long range dependence. Ann. Appl. Probab. **28**(1), 35–78 (2018)
5. Bhattacharjee, M., Banerjee, M., Michailidis, G.: Change point estimation in a dynamic stochastic block model. arXiv preprint arXiv:1812.03090 (2018)
6. Donnat, C., Holmes, S., et al.: Tracking network dynamics: a survey using graph distances. Ann. Appl. Stat. **12**(2), 971–1012 (2018)
7. Doyle, P., Snell, J.: Random walks and electric networks. AMC **10**, 12 (1984)
8. Ellens, W., Spieksma, F., Mieghem, P.V., Jamakovic, A., Kooij, R.: Effective graph resistance. Linear Algebra Appl. **435**(10), 2491 – 2506 (2011). http://www.sciencedirect.com/science/article/pii/S0024379511001443
9. Ghosh, A., Boyd, S., Saberi, A.: Minimizing effective resistance of a graph. SIAM Rev. **50**(1), 37–66 (2008)
10. Ho, Q., Song, L., Xing, E.P.: Evolving cluster mixed-membership blockmodel for time-varying networks. J. Mach. Learn. Res. **15**, 342–350 (2015)
11. Kim, B., Lee, K.H., Xue, L., Niu, X., et al.: A review of dynamic network models with latent variables. Stat. Surv. **12**, 105–135 (2018)
12. Klein, D., Randić, M.: Resistance distance. J. Math. Chem. **12**(1), 81–95 (1993)

13. Koutra, D., Shah, N., Vogelstein, J.T., Gallagher, B., Faloutsos, C.: DELTACON: principled massive-graph similarity function with attribution. ACM Trans. Knowl. Discov. Data (TKDD) **10**(3), 28 (2016)
14. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Soc. (2009)
15. Lyons, R., Peres, Y.: Probability on trees and networks (2005). http://mypage.iu.edu/~rdlyons/
16. Monnig, N.D.: The Resistance-Perturbation-Distance. https://github.com/natemonnig/Resistance-Perturbation-Distance (2016)
17. Monnig, N.D., Meyer, F.G.: The resistance perturbation distance: a metric for the analysis of dynamic networks. Discrete Appl. Math. **236**, 347 – 386 (2018). http://www.sciencedirect.com/science/article/pii/S0166218X17304626
18. Peel, L., Clauset, A.: Detecting change points in the large-scale structure of evolving networks. In: AAAI, pp. 2914–2920 (2015)
19. Sricharan, K., Das, K.: Localizing anomalous changes in time-evolving graphs. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, pp. 1347–1358. ACM (2014)
20. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.F., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B., et al.: High-resolution measurements of face-to-face contact patterns in a primary school. PLoS One **6**(8), e23176 (2011)
21. Sylvester, J.A.: Random walk hitting times and effective resistance in sparsely connected Erdős–Renyi random graphs. arXiv preprint arXiv:1612.00731 (2016)
22. Tang, X., Yang, C.C.: Detecting social media hidden communities using dynamic stochastic blockmodel with temporal Dirichlet process. ACM Trans. Intell. Syst. Technol. (TIST) **5**(2), 36 (2014)
23. Von Luxburg, U., Radl, A., Hein, M.: Hitting and commute times in large random neighborhood graphs. J. Mach. Learn. Res. **15**(1), 1751–1798 (2014)
24. Wills, P.: The NetComp Python library (2019). https://www.github.com/peterewills/netcomp
25. Wills, P., Meyer, F.G.: Change point detection in a dynamic stochastic block model (2019). https://ecee.colorado.edu/~fmeyer/pub/WillsMeyer2019.pdf
26. Wills, P., Meyer, F.G.: Metrics for graph comparison: a practitioner's guide. arXiv preprint arXiv:1904.07414 (2019)
27. Wilson, J.D., Stevens, N.T., Woodall, W.H.: Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model. arXiv preprint arXiv:1605.04049 (2016)
28. Wolfe, P.J., Olhede, S.C.: Nonparametric graphon estimation. arXiv preprint arXiv:1309.5936 (2013)
29. Xing, E.P., Fu, W., Song, L.: A state-space mixed membership blockmodel for dynamic network tomography. Ann. Appl. Stat. **4**(2), 535–566 (2010)
30. Xu, K.: Stochastic block transition models for dynamic networks. In: Artificial Intelligence and Statistics, pp. 1079–1087 (2015)
31. Yang, T., Chi, Y., Zhu, S., Gong, Y., Jin, R.: Detecting communities and their evolutions in dynamic social networks-a Bayesian approach. Mach. Learn. **82**(2), 157–189 (2011)