MULTISCALE ANALYSIS OF TIME SERIES OF GRAPHS

Jason D. Lee[1], *Mauro Maggioni*[2]

[1]Department of Mathematics and Computer Science, Duke University, U.S.A.
[2]Institute for Computational and Mathematical Engineering, Stanford University, U.S.A.
Emails: jdl17@stanford.edu, mauro@math.duke.edu

ABSTRACT

Time series of graphs arise in a variety of applications, from the analysis of network traffic to time-dependent data sets to social networks. We introduce novel techniques for measuring distances between graphs in a multiscale fashion, in such a way that such distances reflect changes at different levels of granularity, are easily localized to regions of the graph, and, when considered at a fixed scale, are robust with respect to changes and finer scales. These distances, or similarity measures, are based on random walks on graphs at multiple time scales. We then employ these notions to analyze time series of graphs, and show that on simple models of graphs that possess a natural multiscale structure, our algorithms have desirable properties of sensitivity and robustness.

Keywords— Multiscale analysis, spectral graph theory, dynamic graphs, random walks.

1. INTRODUCTION

We are interested in developing methods for analyzing a time series of graphs $G^{(t)}$, where t is a time parameter, and $G^{(t)}$ is a weighted graph, which in this paper will also be undirected. Graphs that evolve in time arise in a wide variety of applications: friendship graphs in social networks, citation networks, but also graphs that arise for time-varying data, for example geographic graphs associated to mobile wireless phones, nearest neighbor graphs constructed on point clouds that vary in time, and so on. Such graphs are often very large, and coping with their size means, among other things, that different levels of detail may be interesting, depending on the application or task at hand. For this reason, we are interested in methods that are able to "look at" the graph and its dynamics at different "scales", and detect and quantify changes at different scales. Secondly, these graphs are often "noisy", for example because measurements from which the graphs are constructed are often inexact or noisy.

MM is greteful for support from NSF (DMS 0650413, CCF 0808847, IIS 0803293), ONR N00014-07-1-0625 and the Sloan Foundation. J. Lee was partially supported by the Duke PRUV program and NSF IIS 0803293 and DMS 0650413.

We shall assume in this paper that $G^{(t)}$ and $G^{(t+1)}$ have vertices with IDs that are known, and that a suitably large fraction of vertices of $G^{(t)}$ are in $G^{(t+1)}$, and $G^{(t+1)}$ is not much larger than $G^{(t)}$. In other words only a relatively small number of vertices is added or subtracted from the vertex set of $G^{(t)}$ in order to obtain the vertex set of $G^{(t+1)}$.

The first task is to define *a notion of similarity, or distance, between a pair of graphs*, that can quantify differences between the two graphs at different scales. Such similarity will have some "averaging property", in the sense that if "small", "noisy" perturbations occur at fine scale, they should only induce a small change in the similarity at a coarser scale. These considerations suggest a multiscale approach, where in fact a multiscale family of similarities, at different scales and locations, is to be considered. A global similarity may then be defined, if needed, by an appropriately weighted average of the multiscale similairities. This should be compared to the notion of graph isomorphism, common in Computer Science: this combinatorial notion is not flexible enough for our purposes. For example, if two graphs have different number of vertices, or differ by ϵ on any edge weight, then they are not isomorphic. Moreover, graph Isomorphism also has no known polynomial time algorithm and is not practical to compute for even moderately sized graphs ($1000+$ nodes). In the Social Network Analysis community an opposite viewpoint has emerged, when large scale statistics of the network, such as degree-distribution, clustering coefficient, radius and diameter, have been used to compare networks. These quantities are typically easy to compute and do allow comparison between two graphs of different sizes; however, they are global in nature and not discriminative.

The second task is to apply such similarity measure(s) between $G^{(t)}$ and $G^{(t+1)}$ and track them through time. At this point such similarities become statistics of the network through time and may be analyzed by real-valued time series techniques.

2. MULTISCALE RANDOM WALKS

We construct compressed representations of graphs, in order to be able to detect and quantify changes in

graphs at multiple resolutions. We do this by using random walks at different time scales. Let G be a connected weighted graph with n vertices. The $n \times n$ matrix of weights is denoted by W, and it will be assumed symmetric and nonnegative. We let the degree matrix D be the diagonal matrix defined by $D_{ii} = \sum_j W_{ij}$; we assume without loss of generality that $D_{ii} > 0$. Then the random walk on G [1, 2] is the Markov chain with transition matrix $P = D^{-1}W$: P_{ij} represents the probability of jumping from j when in vertex i, and because of the Markov property, P^t_{ij} is the probability of being at j at time t starting from i. We think of P as a large sparse matrix, since most vertices typically have a small number of edges. Typically P has a decaying spectrum, and when powers of P are taken, the numerical rank of P, i.e. the number of singular values above a fixed threshold, decreases. We may think of the row $P^t_{x,:}$ as a probability distribution "centered" at x and with a certain "variance" increasing with t. The rate of increase of the variance highly depends on the geometry of the graph.

For a set of increasing scales t_j, for example $t_j = 2^j$, we may choose a subset of rows of P^t that spans, up to a certain precision, the whole row space of P^t. In this way we would obtain a compressed representation of the random walk at the scale. This, and numerically stable and efficient algorithms that pursue this multiscale analysis of graphs are described in [3].

Here we choose a simpler route, in order to focus on the main ideas. We construct a hierarchical partition of the graph, i.e. a family of connected subsets $\{\{C_{j,k}\}_{k\in\mathcal{K}_j}\}_{j\geq 0}$ of the graph with the following properties:

(i) for each $j \geq 0$, G is the disjoint union of $\{C_{j,k}\}_{k\in\mathcal{K}_j}$;

(ii) the cardinality of $C_{j,k}$ is roughly $2^{-j}|G|$.

(iii) each $C_{j,k}$ is the disjoint union of $C_{j+1,k'}$, for k' in a subset of $\mathcal{K}_{j+1}$ which we denote $\mathrm{child}(j,k)$. Moreover, every $k' \in \mathcal{K}_{j+1}$ is in one and only one set of children.

There is a natural tree structure associated with the $C_{j,k}$'s, which we denote by $\mathcal{T}$, where each node of the tree is associated with a $C_{j,k}$. If we let $\chi_{j,k}$ be the characteristic function of $C_{j,k}$, defined by $\chi_{j,k}(x) = 1$ for $x \in C_{j,k}$ and 0 otherwise, we may immediately construct a multiresolution analysis for functions on G by defining $V_j := \langle\{\chi_{j,k}\}_{k\in\mathcal{K}_j}\rangle$, so that $V_j \subseteq V_{j+1}$, and thus we have a set of nested closed subspaces of functions on G. We may normalize $\chi_{j,k}$ in various ways which we have no space to discuss here [4]. Moreover, we may let W_{j+1} be the orthogonal complement of V_j into V_{j+1}, and construct an orthonormal basis for W_{j+1}, for example by running the Gramm-Schmidt procedure to orthogonalize $\{\chi_{j+1,k'}\}_{k'\in\mathrm{child}(j,k)}$ to $\chi_{j,k}$, for any $k \in \Gamma_j$, to obtain an orthonormal set $\{\psi_{j+1,k'}\}$ (of cardinality $|\mathrm{child}(j,k)| - 1$), and taking the union of the orthogonal bases thereby obtained. Notice that in this way we obtained an orthonormal basis for W_{j+1}, whose elements are localized in the sense that they come in groups of basis functions whose support is contained in $C_{j,k}$.

For every scale j, and any random walk time scale τ, we may compress P^τ at scale j by restricting P^τ on V_j: this is the operator $[P^\tau]_j$ defined by $\pi_j^*(P^\tau)\pi_j$, where π is the stationary distribution of P, i.e. $\pi P = \pi$. This may be represented by a matrix of size $|\mathcal{K}_j| \times |\mathcal{K}_j|$ whose k_1, k_2 entry is

$$([P^\tau]_j)_{k_1,k_2} := \langle P^\tau \chi_{j,k_1}, \chi_{j,k_2}\rangle \qquad (1)$$

We think of $[P^\tau]_j$ as representing a summary at scale j of the random walk dynamics at time scale τ. Observe that since $V_j \subseteq V_{j+1}$, we may write $\chi_{j,k} = \sum_{k'\in\mathrm{child}(j,k)} \chi_{j+1,k'}$, and therefore

$$\begin{aligned}([P^\tau]_j)_{k_1,k_2} &= \langle P^\tau(\sum_{k'\in\mathrm{child}(j,k_1)} \chi_{j+1,k'}), \sum_{k''\in\mathrm{child}(j,k_2)} \chi_{j+1,k''}\rangle \\ &= \sum_{k'\in\mathrm{child}(j,k_1)} \sum_{k''\in\mathrm{child}(j,k_2)} ([P^\tau]_{j+1})_{k','k''}\ ,\end{aligned}$$

showing that $[P^\tau]_j$ is indeed a coarsening of $[P^\tau]_{j+1}$. The geometric multiscale structure of the $C_{j,k}$ gives the multiresolution analysis V_j as well as a multiresolution representation for the compressed operators $[P^\tau]_j$.

In summary, to a weighted graph G we have associated a multiresolution analysis $\{V_j\}$ and a multiscale family of operators $[P^\tau]_j$ parametrized by two scales: a random walk scale τ and a spatial scale j.

3. COMPARISON BETWEEN GRAPHS

Given two graphs $G^{(1)}, G^{(2)}$, with corresponding multiscale decompositions $(C^{(1)}_{j,k})$ and $(C^{(2)}_{j,k})$ and trees $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$. For the moment being we shall assume that there are injections $\varphi_i : G^{(i)} \to G$, for $i = 1, 2$, such that $\varphi^{(1)}(G^{(1)}) \cap \varphi^{(2)}(G^{(2)})$ is large compared to $|G^{(1)}|$ and $|G^{(2)}|$, and such that there exists a multiscale partition on such intersection that is isomorphic to both $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$. Without loss of generality, we may assume that under these isomorphisms, $C^{(1)}_{j,k}$ corresponds to $C^{(2)}_{j,k}$. We shall assume that this isomorphism is given, or easy to estimate, for example because for each $C^{(1)}_{j,k}$ there exists a unique $C^{(2)}_{j,k}$ with large vertex overlap with $C^{(1)}_{j,k}$.

Now, at every scale we have $[(P^{(i)})^\tau]_j$, for $i = 1, 2$, whose respective (k_1, k_2) elements correspond to the "same" portion of the graphs $C^{(i)}_{j,k_1}$ and $C^{(i)}_{j,k_2}$. We may then define a family of distances between the graphs $G^{(1)}$ and $G^{(2)}$ at scale j by

$$\beta_{j,\tau} := ||[(P^{(1)})^\tau]_j - [(P^{(2)})^\tau]_j||_{\mathbb{F}}\,,$$

were τ is an "internal mixing time" parameter. A distance between $G^{(1)}$ and $G^{(2)}$ by averaging across scales: $d_{\alpha,\tau}(G^{(1)}, G^{(2)}) = \sum_j 2^{-j\alpha} \frac{\beta_{j,\tau}}{|\mathcal{K}_j|}$, for a "smothness" parameter $\alpha \geq 0$. For small values of α, changes at all scales are weighted similarly, while for large values of α changes at finer scales are down-weighted compared to changes at the coarsest scales.

Observe that these measures of distances may be localized, for example $|[(P^{(1)})^\tau]_j(k_1, k_2) - [(P^{(2)})^\tau]_j(k_1, k_2)|$ measures the change at scale j in the transport of the random walk at time τ from C_{j,k_1} to C_{j,k_2}. This also suggests that there is a natural choice of τ for every j, since τ should not be too small so that no mass is transported between two different clusters, nor too large so that the random walk is close to its stationary distribution and the matrices above have approximate rank 1. An acceptable value of τ will be selected according to such a criteria, unless otherwise mentioned; for example we choose τ so that the sum of the diagonal entries of $[(P^{(1)})^\tau]_j$ is not too large, and the numerical rank is larger than 1. Additionally, one may consider the wavelet representation of $(P^{(1)})^\tau$ at scale j: instead of compressing it by projecting onto $\chi_{j,k}$ it could be compressed by projecting onto the $\psi_{j+1,k'}$, but we have no space to discuss this.

4. EXPERIMENTS

We test the multiscale analysis algorithm on a simulated graph time series. The purpose of the simulation is to illustrate that our methodology can adequately detect several types of dynamics, and capture multiscale changes. The first type of dynamics is vertices leaving and entering. Next, the edge weights on the graph will change between each graph.

4.1. Graphs with multiscale structure

We construct families of graphs with a clear multiscale structure. We believe this construction is interesting *per se*, since we could not find in the literature constructions of graphs with a natural multiscale decomposition, except in some very symmetric cases. The construction is as follows: we fix a dictionary of graph types $\{\mathcal{G}_i\}$, each class having possibly a certain number of parameters, for example $\mathcal{G}_1$ may be the class of dyadic trees of a certain depth, $\mathcal{G}_2$ could be the class of complete graphs of a certain order. At the coarsest scale, we let G_1 be an element of the dictionary, and assign to each edge a weight 1. To add one scale to G_1, we pick a second element of the dictionary, assign to each edge a weight $\gamma > 1$ (in the following experiments, $\gamma = 8$), and obtain G_2 be replacing each vertex of G_1 by this new graph. We continue in this fashion, at scale $j+1$ picking an element from the dictionary of graphs, weighting its edges by γ^j, and obtaining G_{j+1} by replacing each vertex of G_j by this new weighted graph. This simple construction may be generalized in several ways to become quite general, for example by allowing, at each scale, a different graph type at every vertex, allowing random weights of order γ^j, etc...This is repeated till scale J. See Figure 1a for a realization of such multiscale graph, with 3 scales.

These graphs have a natural multiscale structure since a random walk started at a vertex at the finest scale J will spend, with high probability, a relatively large time (i.e. number of steps) in the subgraph corresponding to the J-th scale, before exiting this subgraph along an edge from scale $J-1$, and entering into another subgraph corresponding to the J-th scale. The spectrum of P for such graphs is approximately the sum of multiple steps, capturing the different time scales.

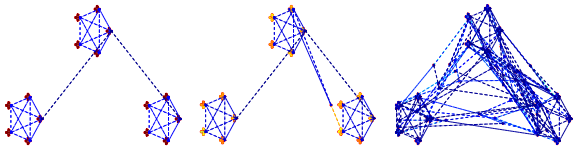
4.2. Stochastic Dynamics: Vertices and Edges

At each time step we introduce new vertices in a random fashion: the first neighbor of a new vertex is chosen at random over the entire graph. The vertex then chooses his edge weights according to a fixed distribution with larger weights assigned a smaller probability. The remaining neighbors are chosen with the distribution $F(v) = \alpha F_{\text{diff}}(v) + (1-\alpha)\text{Unif}(v)$ where F_{diff} is the probability of the introduced vertex of diffusing to another vertex after time τ and $\text{Unif}(v)$ is the uniform distribution over the vertices. We also add edge noise to all the edge weights. The edge noise has distribution $N(0, \sigma_{ij})$ where $\sigma_{ij} = \frac{W(i,j)}{10}$.

We use an online method to determine the cluster assignments of a vertex z that is joining the graph. Let C represent the partition prior to z joining and $C_{z,l}$ represent the partition after vertex z joins C^l. Let $J(C, \tau) = \sum_{i=1}^k P^\tau(C^i, C^i)$. We assign z to the C^l that maximizes $J(C_{z,l}, \tau)$.

4.3. Description of the Experiment

A time series of 51 graphs was constructed with a single vertex and edge noise added after each graph. The initial graph has 3 layers, and 60 nodes. The edge weights on the scales from finest to coarsest are 64, 8, and 1. The initial partitions are fed to the algorithm and we update the partitions using the aforementioned online algorithm. At each step, a new vertex with 3 edges is introduced. The change in the graph time series is computed as the Frobenius norm between $P^{\tau_j}_{C_{j,t},t} - P^{\tau_j}_{C_{j,t-1},t-1}$ for each scale j and time t. The partition $C_{j,t}$ is updated using the online algorithm above since the number of nodes in each $G^{(t)}$ changes. The partition is recomputed by spectral clustering [5] when there is a large change on the coarse scale, since this indicates these partitions no longer resemble good clusters.



(a) Initial Graph (b) Graph after adding one vertex (c) Final Graph

Fig. 1: Multiscale Graphs at different times. The distance between vertices in the layout is inversely related to the weight of the edges, which is represented also by thickness. The stochastic model is biased towards the creation of new edges and vertices.

4.4. Analysis of the Dynamics

In this example, the graph has at least two different scales, of cardinality $3, 15$. We fix the number of clusters at 2 and fix 3 clusters at the coarse scale and 15 clusters at the finer scale. After several perturbations, the resulting graph may no longer have 3 coarse scale clusters or 15 fine scale clusters, but the dynamics can still be effectively measured by fixing the number of scales and clusters.

As stated before, the change in the graph is measured as $||P^{\tau_1}_{C_{1,t},t} - P^{\tau_1}_{C_{1,t-1},t-1}||_F$ and $||P^{\tau_2}_{C_{2,t},t} - P^{\tau_2}_{C_{2,t-1},t-1}||_F$. This is plotted in Figure 3. We update $C_{j,t}$ using the online updates and recompute $C_{j,t}$ when there is a large change on the coarse scale. The plot corresponding to the change in the fine scale is more jittery. The minimum value it attains is near .1. This is because even small changes in the graph are altering the fine scales. However, the plot of change in the coarse scale takes on much lower values. The large peaks in the coarse scale plot are also large peaks in the fine scale plot. On the other hand, many of the peaks in the fine scale plot are either much smaller or disappear in the coarse scale plot. For example, the $7th$ time point causes a large change in the fine scale, but almost no change on the coarse scale (Figure 3). In fact, we see that the difference between $G^{(7)}$ and $G^{(6)}$ is a vertex with edges to two fine scale clusters joined and thus merged two fine scale clusters. This does not affect the coarse scale since the newly introduced edges are all within the same coarse scale cluster. See Figure 2 for details. The largest change is at time point 27. Before time 27, two of the coarse clusters are tightly interconnected, but one cluster is not connected to the other two. The vertex entering at time 27 causes the lone cluster to be connected to the other two. This causes the large change at time 27 that is seen in Figure 3. See the bottom graphs in Figure 2 for the change between $G^{(26)}$ and $G^{(27)}$.

5. CONCLUSION AND FUTURE WORK

We have constructed a novel notion of similarity between graphs by mapping each graph to a vector of operators representing compressed versions of the random walk at multiple scales. We have shown that on

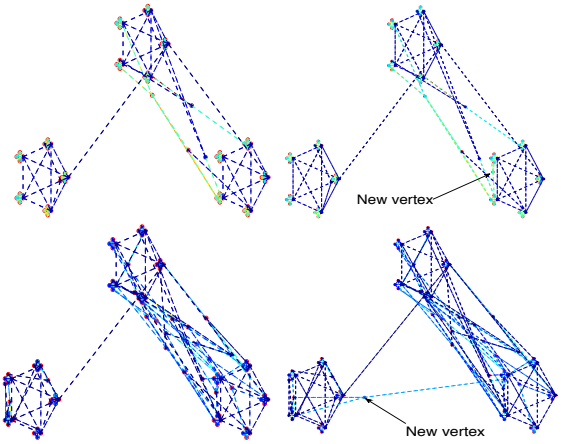


Fig. 2: Top: from $G^{(6)}$ to $G^{(7)}$ a vertex joining causes two fine scale clusters to merge. This is only seen in the fine scale measurements (see Figure 3). Bottom: from $G^{(26)}$ to $G^{(27)}$ a vertex joins two far-away coarse scale clusters. Notice the vertex in the plot on the right, but not in the left side plot. This vertex causes a large change between these two graphs since it joins two coarse scale clusters (see Figure 3).

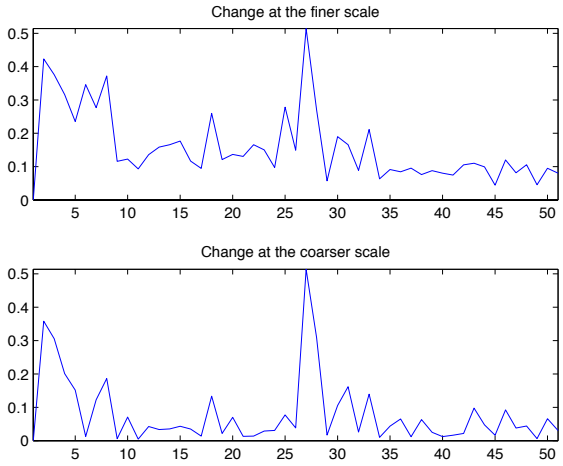


Fig. 3: Plot of the change between $G^{(t)}$ and $G^{(t-1)}$. Y-axis is $||P^{\tau}_{C_{j,t},t} - P^{\tau}_{C_{j,t-1},t-1}||_F$.

toy examples of dynamic graphs with a clear multiscale structure and evolving in time with simple stochastic rules, these multiscale similarity measures capture and quantify changes at different scales.

The above construction may be extended in many directions, in particular in the way the partitions are constructed and updated, the way the parameter τ is selected. Some of these extensions will be discussed in a forthcoming paper [4].

6. REFERENCES

[1] F. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[2] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426–7431, 2005.

[3] R.R. Coifman and M. Maggioni. Diffusion wavelets. *Appl. Comp. Harm. Anal.*, 21(1):53–94, July 2006. (Tech. Rep. YALE/DCS/TR-1303, Yale Univ., Sep. 2004).

[4] J.D. Lee and M. Maggioni. Multiscale dynamic graphs. *in preparation*, 2010.

[5] Ulrike Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.