ANL/MCS-TM-325

Multilevel Method for Modeling Large-Scale Networks

Mathematics and Computer Science Division

About Argonne National Laboratory
Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

Availability of This Report
This report is available, at no cost, at http://www.osti.gov/bridge. It is also available on paper to the U.S. Department of Energy and its contractors, for a processing fee, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone (865) 576-8401
fax (865) 576-5728
reports@adonis.osti.gov

Disclaimer
This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

ANL/MCS-TM-325

Multilevel Method for Modeling Large-Scale Networks

by
Ilya Safro
Mathematics and Computer Science Division, Argonne National Laboratory

June 30, 2011

Abstract

Understanding the behavior of real complex networks is of great theoretical and practical significance. It includes developing accurate artificial models whose topological properties are similar to the real networks, generating the artificial networks at different scales under special conditions, investigating a network dynamics, reconstructing missing data, predicting network response, detecting anomalies and other tasks. Network generation, reconstruction, and prediction of its future topology are central issues of this field. In this project, we address the questions related to the understanding of the network modeling, investigating its structure and properties, and generating artificial networks.

Most of the modern network generation methods are based either on various random graph models (reinforced by a set of properties such as power law distribution of node degrees, graph diameter, and number of triangles) or on the principle of replicating an existing model with elements of randomization such as R-MAT generator and Kronecker product modeling. Hierarchical models operate at different levels of network hierarchy but with the same finest elements of the network. However, in many cases the methods that include randomization and replication elements on the finest relationships between network nodes and modeling that addresses the problem of preserving a set of simplified properties do not fit accurately enough the real networks. Among the unsatisfactory features are numerically inadequate results, non-stability of algorithms on real (artificial) data, that have been tested on artificial (real) data, and incorrect behavior at different scales. One reason is that randomization and replication of existing structures can create *conflicts between fine and coarse scales* of the real network geometry. Moreover, the randomization and satisfying of some attribute at the same time can abolish those topological attributes that have been undefined or hidden from researchers.

We propose to develop multilevel methods to model complex networks. The key point of the proposed strategy is that it will help to preserve part of the *unknown* structural attributes by guaranteeing the similar behavior of the real and artificial model on different scales.

This report is based on the research proposals submitted to DOE [41, 37].

1 Introduction and Motivation

Discovering topological laws of networks has numerous applications in many theoretical and practical studies [12]. For example, the topology of the Internet and many other evolutionary processes can be described by power laws [34, 11, 6], some networks are structured as "bow-tie" or "jellyfish" [11, 46], and small-world model covers a variety of social networks claiming that the average distance between network objects is small [53]. Correct understanding of the network topology and accurate modeling of it lead to the clarification of questions related to the investigation of *abnormal* network behavior, *unusual* evolution of its parts, and *anomalies* detection and, thus, to the prevention of them. These anomalies and their patterns can be interpreted as threats, attacks, viruses, or other events that require a special attention.

Artificial network generation is one of the central tasks of network modeling. One immediate application of an artificial network is in simulation studies and algorithm testing. Another important aspect of motivation is an ability to generate artificial networks that are similar to the real classified (or limited-access) networks for open science research. For example, on one hand, many DOE computer networks (such as in National Labs and OpenScience Grid) is of great interest for

hackers and it is not secure to disclose it. On the other hand, it can be extremely important to test various cybersecurity methods on it by providing a similar artificial networks to the scientific community. An ability to generate an artificial network can provide insight into the real network creation process and can emphasize the presence or absence of common patterns or properties. Most of the known network generation models are based on the randomization or/and the replication of real or/and artificial parts of networks. For example, many *random graph models* are reinforced by some online- or post-processing that preserves a (usually extremely limited) number of predefined properties such as degree distribution [2, 33], maximum number of small cliques, or small diameter. The advantage of random graph models such as *Erdös-Rényi graphs* [21], *Poisson graphs* and randomized power-law models consists of the ability to analyze and to generate them relatively easily at different scales by preserving the same probabilistic laws. The *preferential attachment* modeling extends the simplified random graph modeling by addressing the question of network evolution. The key idea is to identify the probability of connecting a node to one of the most preferred nodes (usually, according to the "rich get richer" principle) [5, 29]. Incorporating *geographical information* is the next step in improving the previous models. The *small-world model* represents networks with high clustering coefficients and low diameter [53]. The analysis of distance between nodes is one of the key interests of this model. The BRITE generator [31] represents this class of models; it combines the geographical Waxman model [54] and the preferential attachment techniques from [5]. *Optimization problem* related models have been developed in order to represent the dynamics and structure of the network that can naturally formulated using some optimization problem. Most of these models focus on a very limited number (usually 1 or 2) of graph properties, ignoring more complex combinations of them and those that are *unknown*. On the other hand, the simplicity of these models has a significant advantage: only a small number of parameters is needed to describe these models and generators. One of the most interesting graph generators developed so far is R-MAT [13], which simultaneously considers multiple properties and addresses a problem at different scales. Another strong replication type of modeling is based on the Kronecker product of graphs [30]; allows one to create networks of different size that preserve the properties of the original network by introducing randomized and original factors. Hierarchical modeling was introduced in [35]. In addition to the scale-free properties, the authors preserve a hierarchical organization of the structure and address the issue of high degree of clustering which is very important in many real-life networks. This approach is most closely related to the proposed method. The main difference is that in [35] the elementary operations are done with finest elements (and their subsets) of the network, while the original geometric relationships between different scales are not modeled. A detailed survey of existing models and generators can be found in [12].

Most of the existing models are formulated by using a simple weighted graph data structure and cannot deal with more complex relationships between network objects. Another common, almost unaddressed issue is the behavior of the model at different scales of coarseness. The R-MAT and Kronecker product models repeat the same property at different parts of the network recursively, essentially dividing an original network into several similarly formulated small networks, renormalizing them, randomizing the connectivity between them, and thus preserving the same

predefined properties over all parts of the network at both micro- and macroscopic parts of the graph. However, these approaches do not address the problem of relationship between *different* macroscopic parts of the model at various resolutions (that may be not connected directly) and the relationships between microscopic and macroscopic scales. Moreover, at different scales these relationships may be more complicated and irregular, may completely change, or even disappear.

The main objective of a *multilevel algorithm* is to create a hierarchy of problems, each representing the original problem but with fewer degrees of freedom. Representation of an original problem at the coarser scales allows to look at the problem at "different resolutions" and then to achieve a qualitative solution by combining the related information from all scales. This method was originally developed for solving elliptic partial differential equations, and up to now, has represented the most effective class of numerical algorithms for them. During the past two decades there were many attempts to employ multilevel strategies for solving combinatorial optimization problems [9, 45, 25, 51, 52, 50, 27, 1]. The multilevel algorithms typically applied to VLSI design [15, 18, 16, 19, 14], graph optimization problems (with a special attention on the partitioning problem [32, 20, 3, 45, 27, 38, 25, 7, 23, 28, 4, 1, 17]), and several others [50, 51, 24, 42, 43, 44, 39, 52]. Most of the multilevel schemes were developed for a *simple graph model*. Whereas the variety of multilevel algorithms for continuous systems turned into a separate field of applied mathematics [8, 10, 48], for combinatorial problems they still have not reached an advanced stage of development, consisting in practice of a very limited number of multilevel techniques. To the best of our knowledge, all existing hypergraph multilevel algorithms are based on the same *strict coarsening* ([47, 49, 26]), which can suffer from serious limitations [44, 9] when the instance of the problem is not very regular. Thus, it inherits all disadvantages of the simple graph model.

The proposed multilevel framework for the large-scale network modeling will address several mathematical and algorithmic challenges:

- How does one coarsen a network attribute?
- How can one design a coarse-to-fine interpolation of a constrained network?
- What are the spectral/algebraic/topological properties of the coarse system?
- How can interactions between network objects be modeled at different scales?
- Given more than one real network of the same nature, how can one design a common artificial model?
- How does one define a coarsening of an anomaly?
- How does one model a multilevel representation of a hypergraph?
- Given a multilevel model, how can one design an online remodeling with small portions of new data?
- How to define a similarity between two objects (nodes, groups of nodes, hyperedges, etc.) in a model?

2 Methods

Most modern network models and their respective generators (if they exist) are based on the set of predefined properties (or attributes) that describe the model. The goal of these generators can be generalized as the following task.

Task 1. *Given an original graph* $G_0 = (V_0, E_0)$ *and its explicit properties* Π*, where* $|V_0| = n_0$ *and* $|E_0| = m_0$*, design a graph* $\tilde{G}_0 = (\tilde{V}_0, \tilde{E}_0)$ *of a size* $(\tilde{n}_0, \tilde{m}_0)$ *(usually* $(\tilde{n}_0, \tilde{m}_0) \neq (n_0, m_0)$*) keeping the same properties* Π*.*

In practice, in many model generators this task can be reformulated as process of designing a *random* graph by using some random model (such as the Erdös-Rényi model) or by accepting *randomized* steps in some iterative approach (that preserves the properties). In more advanced modeling (such as R-MAT [13] or Kronecker product models [30]), the properties are replicated from initial networks and preserved in different geographical parts of the graph. Our central goal consists of introducing a modeling process that (in addition to Π) preserves a geometry and some algebraic properties (explained late) of the original network(s). We expect to achieve this goal by introducing a multilevel modeling process inspired by AMG.

In the multilevel framework a hierarchy of decreasing-size graphs $G_0, G_1, ..., G_k$ is constructed. Starting from the given graph, $G_0 = G$, we *coarsen* it and represent by the sequence $G_1, ..., G_k$, then solve the coarsest level directly, and finally uncoarsen the solution back to G. This process is called a V-cycle. In general, the AMG-based graph coarsening [44] is interpreted as a process of weighted aggregation of the graph nodes to define the nodes of the next-coarser graph. In weighted aggregation each node can be divided into fractions, and different fractions belong to different aggregates. The construction of a coarse graph G_{i+1} from a given G_i is divided into three stages: (1) a subset C of V_i is chosen to serve as the *seeds* of the aggregates (the nodes of the coarse graph); (2) the rules for aggregation are determined, thereby establishing the fraction of each non-seed node belonging to each aggregate; and (3) finally the graph couplings (or edges) between the coarse nodes are calculated. Stages (1) and (2) are finished when the projection fine-to-coarse operator $\uparrow_f^c$ (see Figure 2) for level i is created. In particular, if $X = \{x_{pq}\}$ is a matrix of

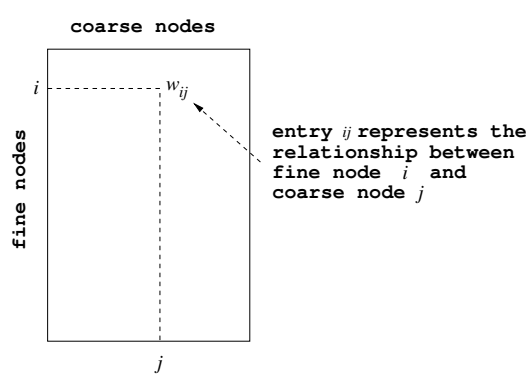

Figure 1: Structure of $\uparrow_f^c$.

off-diagonal entries of the Galerkin fine-to-coarse approximation

$$\uparrow_f^c \cdot L_i \cdot (\uparrow_f^c)^T$$

and $r_p = \sum_{q=1}^{n_k} x_{pq}$ is the sum of its pth row elements, then the Laplacian of coarse graph L_{i+1} is of the form

$$L_{i+1} = X - diag(r_1, r_2, ..., r_{n_{i+1}}) \ . \tag{1}$$

We present, first, an intuitive formulation of the proposed strategy for a multilevel network modeling that can be divided in two tasks.

Task 2. *Given an original graph $G_0 = (V_0, E_0)$ with the corresponding AMG-based hierarchy of Galerkin fine-to-coarse approximations $\mathscr{H} = \{\uparrow_i^{i+1} \cdot L_i \cdot (\uparrow_i^{i+1})^T\}$ and its explicit properties Π, where $|V_0| = n_0$ and $|E_0| = m_0$, design a graph $\tilde{G}_0 = (\tilde{V}_0, \tilde{E}_0)$ of a size $(\tilde{n}_0, \tilde{m}_0)$ satisfying the same properties Π and preserving the hierarchical similarity between $\mathscr{H}$ and $\tilde{\mathscr{H}} = \{\tilde{\uparrow}_i^{i+1} \cdot \tilde{L}_i \cdot (\tilde{\uparrow}_i^{i+1})^T\}$.*

Task 3. *Extend Task 2 for hypergraphs.*

In the remainder of this Section we clarify these tasks by formulating open research directions, mathematical and algorithmic challenges.

2.1 Structure of the V-cycle

Before formulating the questions related to the one- and two-level problems, we explain our viewing of the entire multilevel scheme. We plan to solve the problem of generation by using a bottom-up framework.

In this framework we begin by constructing the multilevel hierarchy $\mathscr{H}$ from the finest to the coarsest level. The coarsest-level network will serve as a a coarsest level for the artificial network, namely, for $\tilde{\mathscr{H}}$. Next, by introducing a new interpolation operator $\tilde{\uparrow}_c^f$ (for an artificial network hierarchy), one can model the next-finest level and similarly the entire $\tilde{\mathscr{H}}$. Throughout both the coarsening and uncoarsening (generation) parts of the V-cycle, set Π will be propagated and modified correspondingly (see Section 2.3) such that at the finest level (at $\tilde{G}_0$) the original Π will be preserved. The scheme is presented in Figure 2.

Challenge 1. *What should be the coarsest level for the multilevel construction? How can one identify it, given the initial Π? How can one identify when some property disappears and the following coarsening is meaningless?*

2.2 Similarity between $\mathscr{H}$ and $\tilde{\mathscr{H}}$

The problem of similarity between $\mathscr{H}$ and $\tilde{\mathscr{H}}$ is one of the central challenges of this project. One can think about many essentially different ways to express it. The goal of such similarity is to be able to compare two-level projections $\tilde{\uparrow}_i^{i+1} \cdot \tilde{L}_i \cdot (\tilde{\uparrow}_i^{i+1})^T$ and $\uparrow_i^{i+1} \cdot L_i \cdot (\uparrow_i^{i+1})^T$ (with possible scaling when $\tilde{\mathscr{H}}$ is much bigger than $\mathscr{H}$).

We propose to introduce of a *scale-dependent* function ε and parameter δ such that

$$||\varepsilon(\uparrow_c^f) - \uparrow_c^f ||_F < \delta \tag{2}$$

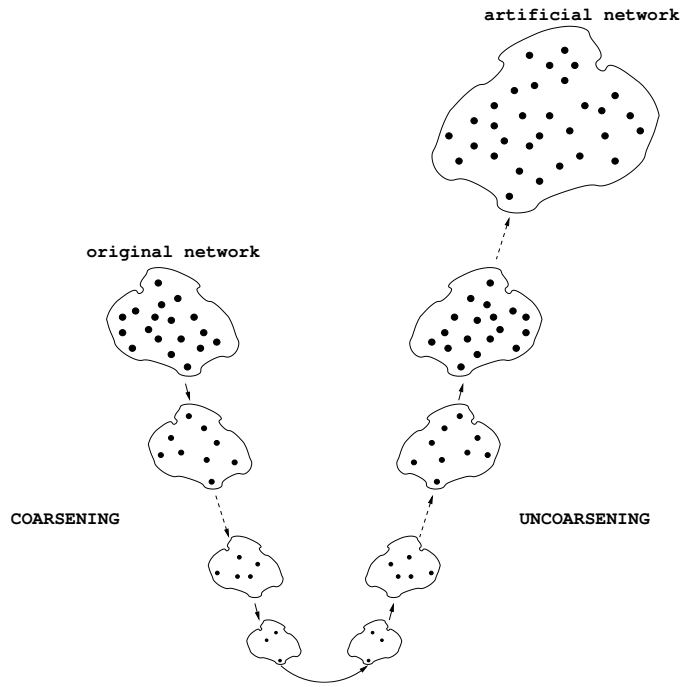

Figure 2: General scheme of the V-cycle.

and then to substitute $\tilde{\uparrow}_c^f$ for $\varepsilon(\uparrow_c^f)$. Since the interpolation operator represents the relationships between fine and coarse points, there are only two ways to change it.

The first way consists of changing the weights of the connections between fine node and its coarse neighborhood. For example, if r_{ij} are the entries of $\uparrow_c^f$, then new entries will be of the form $\alpha_{ij} r_{ij}$ such that $\alpha_{ij} \in (1-\epsilon, 1+\epsilon)$ and for every row i, $\sum_j \alpha_{ij} r_{ij} = 1$. It is easy to see that changing the interpolation weights at level $i+1$ affects the weights at the next-finest levels.

Challenge 2. *How can one find the coefficients α_{ij} at level l, given initial distribution μ of edge weights at level $l-1$? How can this approach be generalized for all levels?*

The second way is to change the structure of the network; in other words, the original and the artificial network matrices will not be pattern isomorphic anymore. One can achieve this by adding more nonzero entries to $\uparrow_c^f$ (or by removing them). This option is more promising and complicated, however, and immediately poses many questions.

Challenge 3. *What are the algebraic properties of $\uparrow_c^f$ and L_c that $\tilde{\uparrow}_c^f$ and $\tilde{L}_c$ must satisfy up to some tolerance?*

In addition to spectral analysis of the Laplacians, we propose to change the structural properties of $\uparrow_c^f$ by bounding the following quantity at level l,

$$\sum_{k=1}^{K} |\tilde{\lambda}_k - \lambda_k| < \beta_l \ , \tag{3}$$

where $\tilde{\lambda}_k$ and λ_k are the lowest eigenvalues of $(\tilde{\uparrow}_c^f)^T \tilde{\uparrow}_c^f$ and $(\uparrow_c^f)^T \uparrow_c^f$, respectively, and K is sufficiently small. This approach opens a problem of finding scale-dependent bounds β_l. We propose

to introduce an *algebraic distance-based restriction model* for introducing new nonzero entries (or for removing the existing ones) for the interpolation operators.

Consider a Jacobi stationary iterative method for solving a linear system $AX = (D+L+U)X = B$, where D, L, and U correspond to the diagonal, lower triangular part, and upper triangular part of A, respectively. Its $(k+1)$th iteration is defined by the following scheme,

$$X_{JAC}^{(k+1)} = D^{-1}(B - (L+U)X_{JAC}^{(k)}) \,, \tag{4}$$

and the corresponding successive overrelaxation (SOR) by

$$X_{NEW\ JAC}^{(k+1)} = (1-\omega)X^{(k)} + \omega X_{JAC}^{(k+1)} \,, \tag{5}$$

where ω is a convergence acceleration parameter. We will use the first t iterations of Jacobi ω-SOR in order to determine the weakly connected edges of the graph. Let A be the graph Laplacian, $B = 0^{n\times r}$, and random matrix $X^{(0)} \in (-\frac{1}{2}, \frac{1}{2})^{n\times r}$.

Definition 1. *Given a graph G and assuming X is an outcome of t iterations of (5), then the algebraic distance between nodes i and j is a function $\rho : V \times V \to (0,1)$ defined by*

$$\rho_{ij} = \max_{l=1}^{r}\{|x_i^l - x_j^l|\} \,.$$

One can prove that if for every i, j, and l, $|x_i^l - x_j^l| \neq 0$, then ρ is a *metric* on the graph. The introduced notion of *algebraic distance* is a generalization based on the principle of obtaining low-residual error components used in the Bootstrap AMG [8]. When a priori knowledge of the nature of this error is not available, slightly relaxed random vectors are used to approximate it. Similarly, we can introduce an algebraic distance based on *Gauss-Seidel* relaxation and *nonsymmetric* relaxations. This approach was successfully introduced in [40] to express the connectivity between different (not necessarily connected by one edge) pairs of nodes in multilevel approaches for combinatorial optimization problems.

We model a new interpolation operator by adding new nonzero entries or changing the existing ones in such way that the algebraic distances will not be strongly affected.

Challenge 4. *How can one minimize the changes of affected ρ_{ij} when having as many changes of r_{ij} as possible?*

The next important property that we propose to preserve when generating new interpolation operator is the Kullback-Leibler (K-L) divergence. The coarse levels are constructed by prolongation of a network information from fine nodes to coarse nodes and thus, this measure can be suitable to establish the limited difference between the original and artificial operators. Given probability distribution of the connections between fine and coarse nodes P and $\tilde{P}$ (in $\uparrow_c^f$ and $\tilde{\uparrow}_c^f$, respectively), the discrete K-L divergence of P from $\tilde{P}$ is defined to be

$$D_{KL}(P||\tilde{P}) = \sum_i P(i) \log \frac{P(i)}{\tilde{P}(i)}. \tag{6}$$

Challenge 5. *What is the best way to bound Kullback-Leibler divergence in order to preserve the similarity between interpolation operators?*

Another possible way to generate $\tilde{\uparrow}_c^f$ is to project $(\uparrow_c^f)^T \uparrow_c^f$ into a lower dimensional space and then to recover from it the corresponding new $(\tilde{\uparrow}_c^f)^T \tilde{\uparrow}_c^f$. However, this method poses more complicated questions of satisfying the previously mentioned properties.

2.3 Coarsening of Properties and Anomalies

The next central question is related to the fulfillment of the properties Π for $\tilde{G}_0$. We propose to begin with two strategies to address this issue. One is to generate an artificial network (that will certainly violate some of the properties at the beginning) and then to apply some (possibly minimalistic) changes to it in order to satisfy Π. Another strategy is to create a multilevel hierarchy $\mathscr{H}_\pi$ for each $\pi \in \Pi$. Clearly, both cases depend on the particular properties that are of interest for concrete applications. The list of *known* properties for the network modeling is well studied, and we will omit it in this paper (for the details see [12]). In this proposal we generalize our ideas by using a general term "property π".

The first strategy consists of purely combinatorial challenges since when the artificial network $\tilde{G}_0$ is already created, the main problem can be formulated as follows.

Challenge 6. *Given a graph $\tilde{G}_0$ and set of properties Π, find a minimum number of legal changes for it, namely either defining (removing) an edge/vertex or changing an edge weight, so that a new graph will satisfy Π.*

In fact, this approach is similar to the existing randomized models-based approaches with the only difference that an initial graph is not empty but is already given and thus some hidden geometrical properties are preserved.

A more complicated and more promising strategy is to create a multilevel hierarchy for Π. In this case, the first problem that can arise is as follows.

Challenge 7. *How can one coarsen the property $\pi \in \Pi$ and coarsen the entire set of properties?*

Coarsening of an anomaly expected to be very similar if not exactly the same as the coarsening of properties. In many cases, the anomalies can be described by special topological structures such as a very dense small subgraph or unusual degrees of nodes. The coarsening of anomalies is expected to allow to detect them at the intermediate coarse levels when their size will be smaller and, thus, the complexity of finding them will be lower.

2.4 Modeling an Artificial Network from Several Original Networks

In many situations the modeling process can be based on several similar instances. For example, different biochemical pathways can form (hyper)graphs of different structure, or different parts of one network can be represented as separate networks. In all these cases, there is a need for analysis and the ability to generate the network with properties that are similar to more than one network.

After coarsening of ν networks, we will get ν different AMG hierarchies $\mathscr{H}_1, \mathscr{H}_2, ..., \mathscr{H}_\nu$. At the first stage, in order to simplify the overall process, this coarsening should be reinforced by introducing a common principle of keeping always the same number of coarse network nodes at the same levels.

Challenge 8. *How can one unify the algebraic/topological properties of several interpolation operators* $\{\uparrow^f_{c,q}\}_{q=1}^{\nu}$ *in one* $\tilde{\uparrow}^f_c$*?*

We propose to begin with the strategy that extends inequality (3) by accumulating the differences for all available ν interpolation operators at scale l,

$$\sum_{q=1}^{\nu}\sum_{k=1}^{K}|\tilde{\lambda}_{k,q} - \lambda_k| < \beta_l \ , \tag{7}$$

where $\tilde{\lambda}_{k,q}$ and λ_k are the lowest eigenvalues of $(\tilde{\uparrow}^f_{c,q})^T\tilde{\uparrow}^f_{c,q}$ and $(\uparrow^f_c)^T \uparrow^f_c$, respectively, and K is sufficiently small. This approach also opens a problem of finding scale-dependent constants β_l.

2.5 Similarity and Connectivity between Two Objects in a Network

The notion of similarity and connectivity between two objects (nodes) in a network (graph or hypergraph) is one of the most important tools in the analysis of network properties. For example, it can be used to address questions of relevancy between objects in recommendation systems (similarity) and to estimate the probabilities of attacking network nodes in a neighborhood of already attacked network node (connectivity).

We propose to focus on the problem of modeling similarity and connectivity using an already defined algebraic distance ρ at different levels. The demonstrated relaxation-based method provides an approach for measuring the connectivity of the neighborhoods between two nodes.

Challenge 9. *What is the explicit quantity that the algebraic distance estimates?*

We propose to begin with two ways of estimating the number of (not) simple paths between two nodes in a graph. One way is to check the relationship between ρ_{ij} and number of not simple paths up to length T that can be calculated by

$$\sum_{l=1}^{T}\frac{1}{l^p}L^l \ , \tag{8}$$

where L is a graph Laplacian and p is a penalty factor for the path length. A more complicated and probably more correct way is to compare ρ with the the number of simple paths between two nodes [36] with similar penalty coefficients. Another way is to compare the algebraic distance with the estimation the average commute time of a random walk [22]. We introduced the algebraic distance-based coarsening in [40] and applied for several combinatorial optimization problems. To provide an insight into this method, we will demonstrate a simple example.

Consider a mesh graph with one additional diagonal ij as depicted in Figure 3. All the edges except the diagonal have weight 1; the diagonal has 2. Let the black small circles represent the

nodes of connectivity interest; in other words, we need to model the connectivity between them and i. In spite of the heavier weight of ij in many networks, this connection is not stronger than the connection between i and its closer neighbors. The presence of such *global* connection can create a local conflict in many optimization problems and models. The concept of algebraic distance is created to prevent similar conflicts. We present the algebraic distances for this graph in Table 1.

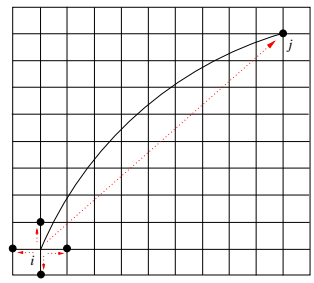

Figure 3: A mesh graph with an additional edge between nodes i and j. The black dots mark some of the nodes selected to serve as the seeds of the coarse aggregates.

The number of Jacobi relaxation sweeps (r) in the experiments varies from 10 to 100 as shown in the leftmost column. Each of the three columns to the right present the ratio between ρ_{ij} and the average of all $\rho_{uv \neq ij}$, for initial graph couplings $w_{ij} = 1, 2, 3$. Clearly the graph coupling between i and j is *decreased* when measured by the algebraic distance. For instance, if the graph coupling between i and j is 1 (as are all other couplings in the graph), then after 10 relaxation sweeps ρ_{ij} is twice as big as the average of $\rho_{uv \neq ij}$, which means that the algebraic coupling between i and j is basically *half* the average coupling over the mesh.

With the algebraic distance we expect to provide a method for modeling the similarity and connectivity for network objects.

Challenge 10. *How one can define a notion of similarity between two graphs by comparing all pair algebraic distances for them?*

Challenge 11. *How can the algebraic distance be extended for hypergraphs?*

r	1	2	3
10	1.98	1.45	1.18
20	2.92	1.81	1.29
50	6.08	3.44	2.38
100	9.87	5.3	3.58

Table 1: Numerical results for the example in Figure 3. The leftmost column shows the number of relaxation sweeps r performed for calculating the algebraic distance ρ. Columns 1-3 correspond to the respective graph weights of the edge ij. All other edges are of equal weight 1. The number of random vectors is 1. All the values are ratios between ρ_{ij} and the average over $\rho_{uv \neq ij}$.

2.6 Online Remodeling of a Network

Often, large-scale networks demand a significant amount of computational resources for the analysis even for the linear time algorithms. If such a network grows dynamically and the newly arrived information can influence the properties of interest and, in general, the network structure, there is a need to be able to remodel the network online and to recompute its properties.

The issue of online network remodeling can be solved naturally within the multilevel framework. If all levels of the multilevel model have been saved and can be accessed, then it is relatively easy to follow the way that a new vertex must pass in order to be added to the existing hierarchy. The amount of the affected coarse aggregates at each level is very limited, and their number can be fairly bounded by $O(\text{number of levels in } \mathscr{H})$. Moreover, when there is an indication that the reinforcement of the existing aggregate with a new vertex is negligible, the adding process can be stopped at the next-coarser level.

While having a recalculated multilevel hierarchy for the original network, both the generation and optimization problems can be finalized with the uncoarsening part only.

3 Conclusions

We proposed to develop multilevel methods to model and generate complex networks. The key point of the proposed strategy is that it will help to preserve part of the *unknown* structural attributes by guaranteeing the similar behavior of the real and artificial models at different scales.

References

[1] Amine Abou-Rjeili and George Karypis. Multilevel algorithms for partitioning power-law graphs. In *IPDPS*, 2006.

[2] Reka Albert and Albert L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[3] Charles J. Alpert, Jen-Hsin Huang, and Andrew B. Kahng. Multilevel circuit partitioning. In *Design Automation Conference*, pages 530–533, 1997.

[4] R. Banos, C. Gil, J. Ortega, and F.G. Montoya. Multilevel heuristic algorithm for graph partitioning. In *Applications of Evolutionary Computing*, pages 143–153, 2003.

[5] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[6] Albert-László Barabási. *Linked: The New Science of Networks*. Perseus Books Group, May 2002.

[7] Stephen T. Barnard and Horst D. Simon. A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. *Concurrency: Practice and Experience*, 6:101–107, 1994.

[8] A. Brandt. Multiscale scientific computation: Review 2001. In T. Barth, R. Haimes, and T. Chan, editors, *Multiscale and Multiresolution methods (Proceeding of the Yosemite Educational Symposium, October 2000)*. Springer-Verlag, 2001.

[9] A. Brandt and D. Ron. Chapter 1 : Multigrid solvers and multilevel optimization strategies. In J. Cong and J. R. Shinnerl, editors, *Multilevel Optimization and VLSICAD*. Kluwer, 2003.

[10] William L. Briggs, Van Emden Henson, and Steve F. McCormick. *A multigrid tutorial: second edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[11] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web: experiments and models. In *Proceedings of the Ninth International World-Wide Web Conference (WWW9, Amsterdam, May 15 - 19, 2000 - Best Paper)*. Foretec Seminars, Inc. (of CD-ROM), Reston, VA, 2000.

[12] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Comput. Surv.*, 38(1):2, 2006.

[13] Deepayan Chakrabarti, Yiping Zhan, and Christos Faloutsos. R-mat: A recursive model for graph mining. In *Fourth SIAM International Conference on Data Mining*, April 2004.

[14] Tony F. Chan, Jason Cong, Tianming Kong, and Joseph R. Shinnerl. Multilevel optimization for large-scale circuit placement. In *ICCAD '00: Proceedings of the 2000 IEEE/ACM international conference on Computer-aided design*, pages 171–176, Piscataway, NJ, USA, 2000. IEEE Press.

[15] Tony F. Chan, Jason Cong, Michail Romesis, Joseph R. Shinnerl, Kenton Sze, and Min Xie. mpl6: a robust multilevel mixed-size placement engine. In *ISPD*, pages 227–229, 2005.

[16] C. Chang, J. Cong, D. Pan, and X. Yuan. Multilevel global placement with congestion control. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 22:395–409, 2003.

[17] Cédric Chevalier and Ilya Safro. Comparison of coarsening schemes for multilevel graph partitioning. In *Proceedings of the 4th Learning and Intelligent Optimization conference*, 2009.

[18] J. Cong and J. R. Shinnerl, editors. *Multilevel Optimization and VLSICAD*. Kluwer, 2003.

[19] Jason Cong, Jie Fang, Min Xie, and Yan Zhang. Mars-a multilevel full-chip gridless routing system. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 24(3):382–394, 2005.

[20] Jason Cong and M'Lissa Smith. A parallel bottom-up clustering algorithm with applications to circuit partitioning in VLSI design. In *Design Automation Conference*, pages 755–760, 1993.

[21] P. Erdos and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[22] Francois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):355–369, 2007.

[23] Bruce Hendrickson and Robert W. Leland. A multi-level algorithm for partitioning graphs. In *Supercomputing*, 1995.

[24] Y. F. Hu and J. A. Scott. A multilevel algorithm for wavefront reduction. *SIAM J. Sci. Comput.*, 23(4):1352–1375, 2001.

[25] G. Karypis and V. Kumar. hmetis 1.5: A hypergraph partitioning package. *Tech. Report, Dept. of Computer Science, Univ. of Minnesota*, 1998.

[26] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. Multilevel hypergraph partitioning: application in vlsi domain. In *DAC '97: Proceedings of the 34th annual conference on Design automation*, pages 526–529, New York, NY, USA, 1997. ACM Press.

[27] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. Multilevel hypergraph partitioning: applications in vlsi domain. *IEEE Trans. Very Large Scale Integr. Syst.*, 7(1):69–79, 1999.

[28] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. Technical Report TR 95-035, 1995.

[29] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63(066123), 2001.

[30] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Gharamani. Kronecker graphs: an approach to modeling networks. Dec 2008.

[31] Alberto Medina, Ibrahim Matta, and John Byers. On the origin of power laws in internet topologies. Technical report, Boston, MA, USA, 2000.

[32] Henning Meyerhenke, Burkhard Monien, and Thomas Sauerwald. A new diffusion-based multilevel algorithm for computing graph partitions of very high quality. In *Proc. 22nd International Parallel and Distributed Processing Symposium, (IPDPS'08)*. IEEE Computer Society, 2008. Best Algorithms Paper Award.

[33] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[34] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[35] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67(2):026112, Feb 2003.

[36] Ben Roberts and Dirk P. Kroese. Estimating the number of s-t paths in a graph. *J. Graph Algorithms Appl.*, 11(1):195–214, 2007.

[37] O. Roderick, I. Safro, and C. Swietlik. Multiscale modeling and generation of data represented by a network. submitted in responde to ANL LDRD calls, Argonne National Laboratory, 2010–2011.

[38] D. Ron, S. Wishko-Stern, and A. Brandt. An algebraic multigrid based algorithm for bisectioning general graphs. Technical Report MCS05-01, Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, 2005.

[39] Dorit Ron, Ilya Safro, and Achi Brandt. A fast multigrid algorithm for energy minimization under planar density constraints. *submitted, Preprint ANL/MCS-P1560-1108*, 2008.

[40] Dorit Ron, Ilya Safro, and Achi Brandt. Relaxation based coarsening scheme for combinatorial optimization problems. *submitted*, 2009.

[41] I. Safro. Multilevel methods for modeling large-scale networks. submitted in responde to DOE Office of Science Program Announcement LAB 09-23, Argonne National Laboratory, 2009.

[42] I. Safro, D. Ron, and A. Brandt. Graph minimum linear arrangement by multilevel weighted edge contractions. *Journal of Algorithms*, 60(1):24–41, 2006.

[43] I. Safro, D. Ron, and A. Brandt. Multilevel algorithm for the minimum 2-sum problem. *Journal of Graph Algorithms and Applications*, 10(2):237–258, 2006.

[44] I. Safro, D. Ron, and A. Brandt. Multilevel algorithms for linear ordering problems. *Journal of Experimental Algorithmics*, 13:1.4–1.20, 2008.

[45] Kirk Schloegel, George Karypis, and Vipin Kumar. Parallel multilevel algorithms for multi-constraint graph partitioning. In *Euro-Par*, pages 296–310, 2000.

[46] S.L. Tauro, C. Palmer, G. Siganos, and M. Faloutsos. A simple conceptual model for the internet topology. 2001.

[47] Aleksandar Trifunovic and William J. Knottenbelt. Parallel multilevel algorithms for hypergraph partitioning. *Journal of Parallel and Distributed Computing*, 68(5):563 – 581, 2008.

[48] Ulrich Trottenberg and Anton Schuller. *Multigrid*. Academic Press, Inc., Orlando, FL, USA, 2001.

[49] C. Aykanat U.V. Catalyurek. Patoh: Partitioning tool for hypergraphs, version 3.0, 2001.

[50] C. Walshaw. A multilevel approach to the travelling salesman problem. Tech. Rep. 00/IM/63, Comp. Math. Sci., Univ. Greenwich, London SE10 9LS, UK, August 2000.

[51] C. Walshaw. A Multilevel Approach to the Graph Colouring Problem. Tech. Rep. 01/IM/69, Comp. Math. Sci., Univ. Greenwich, London SE10 9LS, UK, May 2001.

[52] C. Walshaw. Multilevel refinement for combinatorial optimisation problems. *Annals Oper. Res.*, 131:325–372, 2004.

[53] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[54] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, 1988.

Mathematics and Computer Science Division
Argonne National Laboratory
9700 South Cass Avenue, Bldg. 240
Argonne, IL 60439-4847

www.anl.gov

Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC