Networks are everywhere

Networks are recognized as the standard tool to model complex interconnected systems.

Physical Networks: Power, water, fracture, communication networks

Functional Networks: Supply chains, chemical reaction networks, regulatory networks

Interaction Networks: Cybersecurity, social networks, epidemics

Graphs generated by these models resemble real-world graphs. The Blocked Two-Level (BTER) Graph Model

Core idea of BTER

High clustering coefficients for low degree vertices imply fairly dense clusters. High clustering coefficients for low degree vertices imply many small communities. We explicitly account for this factor. Skewness degree distributions leave enough edges after building these dense blocks to satisfy small world property.

BTER Parameters

Degree distribution or a description of it (e.g., the power-law coefficient) guides the generation. Density parameter controls the density of the smallest blocks. Density decay parameter controls how fast the densities of the blocks decrease with increasing block sizes.

In-depth study of stochastic Kronecker graphs

Stochastic Kronecker Graph (SKG) has been chosen to generate graphs for the Graph500 supercomputer benchmark. It is favored for small number of parameters. Stochastic Kronecker Graph (SKG) has been chosen to generate graphs for the Graph500 supercomputer benchmark. It is favored for small number of parameters. Small-world diameter Many real graphs have amazingly short distances between most pairs of vertices. Scalability and parallelizability We need to generate extremely large instances in an efficient way.

Our goal is to design models that can describe a graph with a small number of parameters. Such models will be instrumental for:

- Insights into
- Generative process
- Graph properties (e.g., eigenvalue distribution)
- Evolution
- Design and analysis of algorithms and architectures
- Alternative to worst-case analysis
- Rigorous studies of heuristics
- Runtime analysis of algorithms
- Benchmarking computers
- Comparing graphs
- Sharing of realistic but non-sensitive data
- Statistically significant graph mining
- Model validation
- Network inference

Graph500 benchmark has been modified to use the noisy SKG model that we have proposed. Our tools are also being used to design future benchmarks.

Requirements for a good model

- Flexibility in degree distribution
- There is no single distribution that works for all graphs, thus a good model should be able to generate graphs with a variety of degree distributions.
- Communities and high clustering coefficients Graphs are known to have many small communities and high clustering coefficients for vertices of all degrees. (i.e., if \(\langle u \rangle \) and \(\langle w \rangle \) are edges, probability of \((u,w)\) should be high). This is a major shortcoming of all scalable graph models.

Small-world diameter Many real graphs have amazingly short distances between most pairs of vertices.

Scalability and parallelizability We need to generate extremely large instances in an efficient way.

Our theoretical and empirical showed SKG cannot generate power-law or lognormal degree distributions by a properly adding noise.

We proved that SKG cannot generate power-law or lognormal degree distributions by a properly adding noise.

We improved the degree distributions by a properly adding noise. We theoretically and empirically showed that this method yields lognormal degree distributions.

- SKG creates notably fewer vertices than intended. For Graph500 parameters 50-75% of the vertices are isolated. We developed techniques to estimate this number for given model parameters, and predict which vertices will be isolated.

Core sizes of SKG graphs are significantly lower than those of the graphs they model, which is a sign of the poor community structure.

In-depth study of stochastic Kronecker graphs

Stochastic Kronecker Graph (SKG) has been chosen to generate graphs for the Graph500 supercomputer benchmark. It is favored for small number of parameters, ease of implementation, full parallelism, and the assumption that graphs generated by these models resemble real-world graphs.

Our analysis of this model provided several important results:

- We proved that SKG cannot generate power-law or lognormal degree distributions, because its degree distribution oscillates between a lognormal and an exponential.

- We improved the degree distributions by a properly adding noise. We theoretically and empirically showed that this method yields lognormal degree distributions.

- SKG creates notably fewer vertices than intended. For Graph500 parameters 50-75% of the vertices are isolated. We developed techniques to estimate this number for given model parameters, and predict which vertices will be isolated.

- Core sizes of SKG graphs are significantly lower than those of the graphs they model, which is a sign of the poor community structure.

In-depth study of stochastic Kronecker graphs

Stochastic Kronecker Graph (SKG) has been chosen to generate graphs for the Graph500 supercomputer benchmark. It is favored for small number of parameters, ease of implementation, full parallelism, and the assumption that graphs generated by these models resemble real-world graphs.

Our analysis of this model provided several important results:

- We proved that SKG cannot generate power-law or lognormal degree distributions, because its degree distribution oscillates between a lognormal and an exponential.

- We improved the degree distributions by a properly adding noise. We theoretically and empirically showed that this method yields lognormal degree distributions.

- SKG creates notably fewer vertices than intended. For Graph500 parameters 50-75% of the vertices are isolated. We developed techniques to estimate this number for given model parameters, and predict which vertices will be isolated.

- Core sizes of SKG graphs are significantly lower than those of the graphs they model, which is a sign of the poor community structure.