Mathematical Techniques for Modeling and Analyzing Large Graphs

Charalampos E. Tsourakakis
charalampos.tsourakakis@aalto.fi

Aalto University

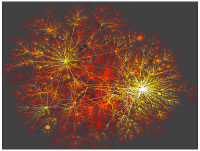
January 14, 2014

Outline

- Age of Networks
  - Types of networks and problems
  - Common characteristics
- Models and Empirical studies
  - Random graphs
  - Triangle counting
- Connectivity
  - Cut structure
  - *Rainbow connectivity*
- Open problems

Age of networks

- Our lives are surrounded by networks.
  - Social networks
  - Technological networks
  - Knowledge and information networks
  - Biological networks



Internet Map

Airline Networks

Online social networks and social media

Networks as Graphs

- Entities – set of vertices
- Pairwise relations among vertices – set of edges
- Can add directions, weights,...
- Graphs model networks
  - Social networks: friendship, collaboration, phone-call networks
  - Technological networks: the internet, power grids, transportation networks
  - Information networks: the World Wide Web, blog networks
  - Biological networks: gene co-expression networks, brain network

Daniel Spielman: "Graph theory is the new Calculus"

Human Brain

Empirical properties of real-world networks

Diverse collections of graphs arising from different phenomena

Are there any typical patterns?

- Static networks
  1. heavy tailed degree sequences
  2. triangles
  3. small-worlds
  4. communities
- Time-evolving networks
  1. densification
  2. shrinking diameters
- The Web graph
  1. bow-tie structure
  2. bipartite cliques

Heavy tails

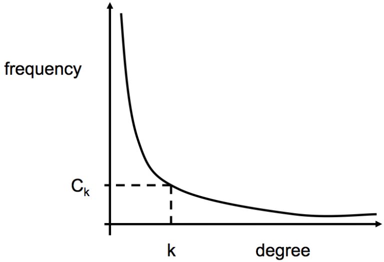
What do the proteins in our bodies, the Internet, a cool collection of atoms and sexual networks have in common? One man thinks he has the answer and it is going to transform the way we view the world.

Scientist 2002

Albert-László Barabási

Degree distribution

- C_k = number of vertices with degree k



- problem : find the probability distribution that fits best the observed data

Power-law degree distribution

- C_k = number of vertices with degree k, then

$$C_k = ck^{-\gamma}$$

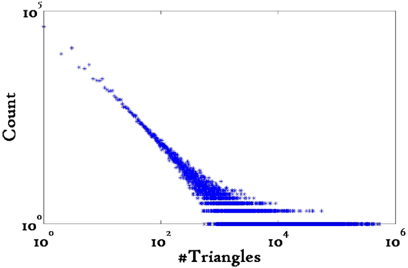
with $\gamma > 1$, or

$$\ln C_k = \ln c - \gamma \ln k$$

- plotting $\ln C_k$ versus $\ln k$ gives a straight line with slope $-\gamma$
- heavy-tail distribution : there is a non-negligible fraction of nodes that has very high degree (hubs)

Heavy tails, triangles

 Triangle



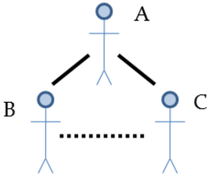
- triangle distribution in `flickr`
- figure shows the count of nodes with k triangles vs. k in log-log scale
- again, heavy tails emerge [Tsourakakis, 2008]

Clustering coefficients

- a proposed measure to capture local clustering is the graph transitivity

$$T(G) = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}$$

- captures "transitivity of clustering"



if u is connected to v and v is connected to w, it is also likely that u is connected to w

Clustering coefficients

- alternative, related definitions
- local clustering coefficient

$$C_i = \frac{\text{Number of triangles connected to vertex } i}{\text{Number of triples centered at vertex } i}$$

- global clustering coefficient

$$C(G) = \frac{1}{n} \sum_i C_i$$

Problems

- the world is full with networks
- what do we do with them?
  - understand their structure (e.g., paths and connectivity, cut structure)
  - understand processes (e.g., random walks, information cascades, epidemics)
  - study their evolution and dynamics
  - create realistic models
  - create algorithms that make use of the network structure

Problems

Algorithms need to scale to large graphs.

- We need to handle datasets with billions of vertices and edges
  - `Facebook`: $\sim$ 1 billion users with avg degree 130
  - `Twitter`: $\geq$ 1.5 billion social relations
  - `Google`: web graph more than a trillion edges (2011)
- We need algorithms for dynamic graph datasets
  - real-time story identification using `twitter` posts
  - election trends, `twitter` as election barometer

Problems: A Billion $ example

Larry Page and Sergey Brin asked...

how to rank Web pages using the network structure?

Pagerank algorithm

Models and Empirical Studies

Random graphs

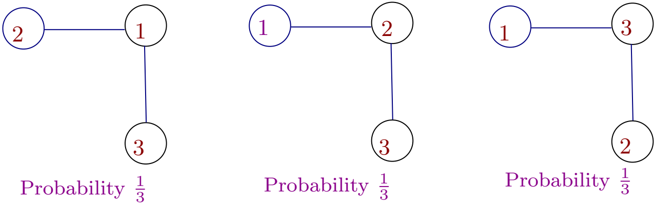
- Erdős-Rényi
- Preferential Attachment
- Random Apollonian networks

Empirical Studies

- Triangle counting

Random graphs

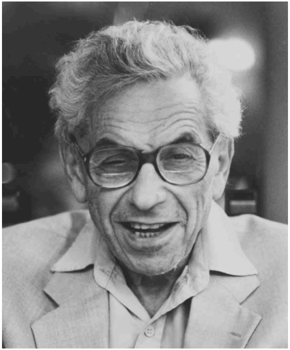
- a random graph is a set of graphs together with a probability distribution on that set
- example



a random graph on $\{1, 2, 3\}$ with 2 edges with the uniform distribution

Random graphs

- Erdős-Rényi (or Gilbert-Erdős-Rényi) random graph model



Paul Erdős
1913 – 1996

Alfréd Rényi
1921 – 1970

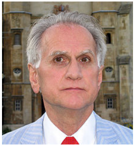
Random graphs

- the $G(n, p)$ model:
- n : the number of vertices
- $0 \leq p \leq 1$: probability
- for each pair (u, v), independently generate the edge (u, v) with probability p
- $G(n, p)$ a family of graphs, in which a graph with m edges appears with probability $p^m(1-p)^{\binom{n}{2}-m}$. Equivalently, $\mathbf{Pr}[G] \propto \left(\frac{p}{1-p}\right)^m$.
- the $G(n, m)$ model: exactly m random edges. Not identical, but related.

Preferential attachment

R. Albert

L. Barabási



B. Bollobás

O. Riordan

growth model:

- at time n, vertex n is added to the graph
- one edge is attached to the new vertex
- the other vertex is selected at random with probability proportional to its degree
- obtain a sequence of graphs $\{G_1^{(n)}\}$.

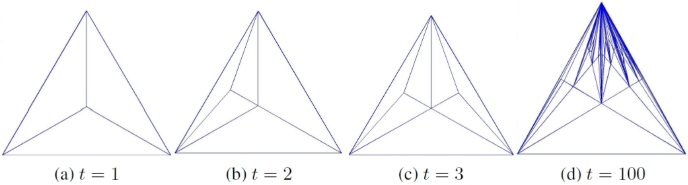
Preferential attachment — generalizations

Barabási-Albert/Bollobás-Riordan preferential attachment generates graph with power law slope equal to 3. Various generalizations exist

- Rather than adding a single edge, we add m edges [Bollobás and Riordan, 2003].
- Tune the power law slope [Buckley and Osthus, 2004].
- Increase clustering coefficients [Holme and Kim, 2002, Saramäki and Kaski, 2004]
- Fitness values, ...

Random Apollonian networks

Are there power-law planar graphs? Yes, [Andrade Jr et al., 2005]!



snapshots of a random Apollonian network (RAN) at:
(a) $t = 1$ (b) $t = 2$ (c) $t = 3$ (d) $t = 100$

- at time $t + 1$ we choose a face F uniformly at random among the faces of G_t
- let (i, j, k) be the vertices of F
- we add a new vertex inside F and we connect it to i, j, k

Random Apollonian networks

- Barabási-Albert/Bollobás-Riordan preferential attachment results in an *expander graph*.
- Real world networks typically have small separators, [Fortunato, 2010].
- Planar graphs have small separators [Planar Separator Theorem, [Lipton and Tarjan, 1979]].
- Also, planar graphs model a wide variety of real world networks including power grids, water distribution and road networks.

Random Apollonian networks

Theorem ([Frieze and Tsourakakis, 2013])

Let $Z_k(t)$ *denote the number of vertices of degree* k *at time* t, $k \geq 3$. *For any* $t \geq 1$ *and any* $k \geq 3$ *there exists a constant* b_k[a] *depending on* k *such that*

$$|\mathbb{E}\left[Z_k(t)\right] - b_k t| \leq K, \quad \textit{where} \quad K = 3.6.$$

Furthermore, for t *sufficiently large and any* $\lambda > 0$

$$\mathbf{Pr}\left[|Z_k(t) - \mathbb{E}\left[Z_k(t)\right]| \geq \lambda\right] \leq e^{-\frac{\lambda^2}{72t}}$$

[a] $b_3 = \frac{2}{5}, b_4 = \frac{1}{5}, b_5 = \frac{4}{35}$ and for $k \geq 6$ $b_k = \frac{24}{k(k+1)(k+2)}$.

Random Apollonian networks

Corollary

For all possible degrees k

$$\mathbf{Pr}\left[|Z_k(t) - \mathbb{E}\left[Z_k(t)\right]| \geq 10\sqrt{t \log t}\right] = o(1).$$

Theorem ([Frieze and Tsourakakis, 2013])

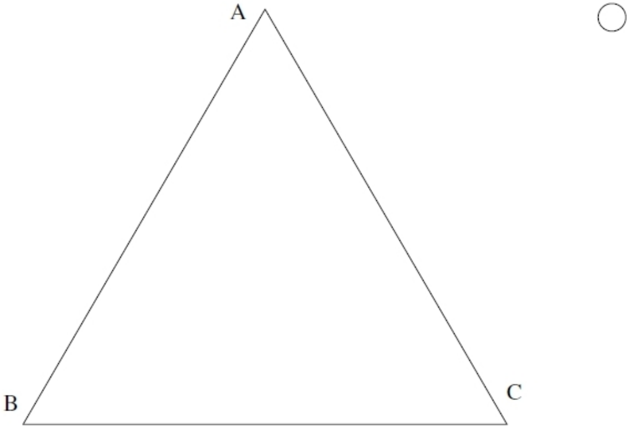
The diameter $d(G_t)$ *of* G_t *satisfies asymptotically* **whp** [a]

$$\mathbf{Pr}\left[d(G_t) > 7.1 \log t\right] \to 0$$

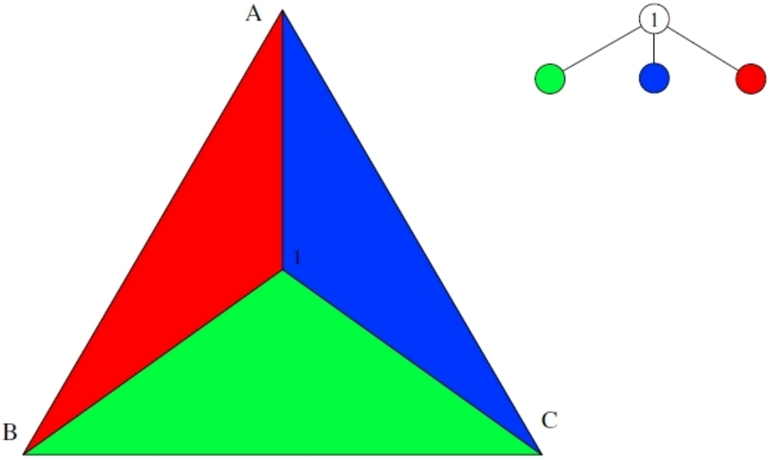
[a] An event A_t holds with high probability (**whp**) if $\lim_{t \to +\infty} \mathbf{Pr}\left[A_t\right] = 1$.

Random Apollonian networks

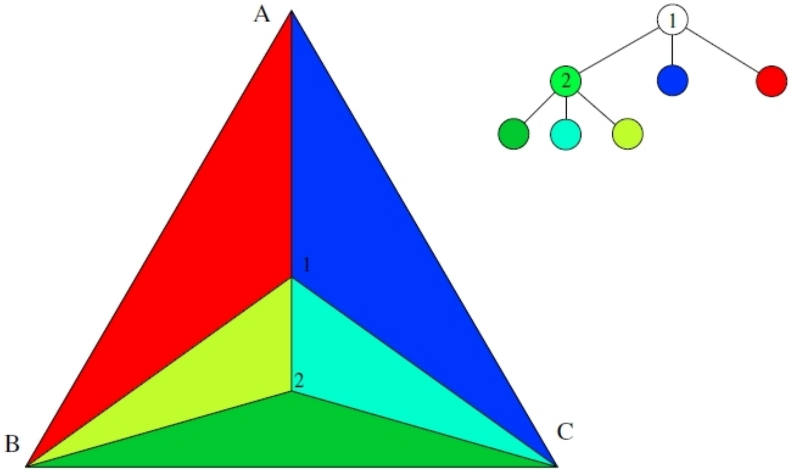
key idea: establish a bijection with random ternary trees



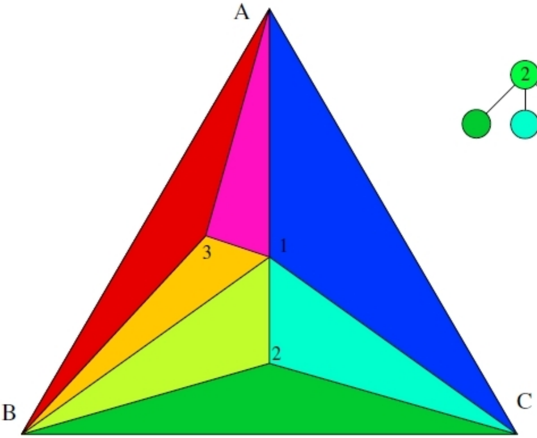
Random Apollonian networks

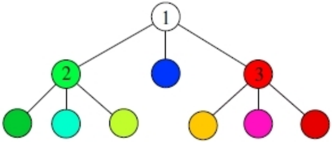


Random Apollonian networks



Random Apollonian networks





Random Apollonian networks

Theorem ([Frieze and Tsourakakis, 2013])

Let $\Delta_1 \geq \Delta_2 \geq \ldots \geq \Delta_k$ be the k highest degrees of the RAN G_t at time t where k is a fixed positive integer. Also, let $f(t)$ be a function such that $f(t) \to +\infty$ as $t \to +\infty$. Then **whp**

$$\frac{t^{1/2}}{f(t)} \leq \Delta_i \leq t^{1/2} f(t)$$

Theorem ([Frieze and Tsourakakis, 2013])

Let k be a fixed positive integer. Also, let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ be the largest k eigenvalues of the adjacency matrix of G_t. Then **whp** $\lambda_i = (1 \pm o(1))\sqrt{\Delta_i}$.

Triangle counting

Numerous applications involve triangle counting, see [Tsourakakis, 2013] and references therein.

- Exponential random graphs generalize Erdős-Rényi random graphs, e.g., $\mathbf{Pr}[G] \propto x_1^m x_2^t$, where t is the number of triangles, $x_1, x_2 > 0$ positive parameters.
- Clustering coefficients and transitivity of a graph
- Uncovering hidden thematic structure in the Web
- Spam detection
- Link recommendation
- Motif detection in biological networks
- Structural balance and status theory
- Microscopic evolution of networks
- Many more!

Triangle sparsifiers

[Tsourakakis et al., 2011]

- Start with graph $G([n], E)$
- Use sparsification parameter p
- Pick a random subset E' of edges such that the events $\{e \in E'\}$, for all $e \in E$ are independent and the probability of each is equal to p.
- $t' \leftarrow$ count triangles on the graph $G'([n], E')$
- Return $T \leftarrow \frac{t'}{p^3}$

How small can p be?

Suppose G is an undirected graph with n vertices, m edges and t triangles. Let also Δ denote the size of the largest collection of triangles with a common edge.

Triangle sparsifiers

Theorem ([Tsourakakis et al., 2011])

Suppose that $\gamma > 0$ is a constant and

$$\frac{pt}{\Delta} \geq \log^{6+\gamma} n, \quad \text{if } p^2\Delta \geq 1, \tag{1}$$

and

$$p^3 t \geq \log^{6+\gamma} n, \quad \text{if } p^2\Delta < 1. \tag{2}$$

for $n \geq n_0$ sufficiently large. Then

$$\mathbf{Pr}\left[|T - \mathbb{E}[T]| \geq \epsilon\mathbb{E}[T]\right] \leq n^{-K}$$

for any constants $K, \epsilon > 0$ and all large enough n (depending on K, ϵ and n_0).

Triangle sparsifiers

- Complexity Analysis: The expected running time of edge sampling is $O(pm)$. If we count in G' using a standard listing triangle algorithm the expected running time is $O(pm + p^2 \sum_i d_i^2)$.
- Expected Speedup is p^{-2}.
- Example: For a graph G with $t \geq n^{3/2}$ and $\Delta \sim n$, we get $p = n^{-1/2}$ implying $O(n)$ expected speedup.
- In practice: Strongly concentrated estimates with an avg. $\sim 10\,000\times$ speedup ($p = 0.01$).

Colorful triangle counting

Can we tighten the analysis and get a smaller p?

No! We need a different algorithm.

The following algorithm achieves optimal performance under no further assumptions on the graph. Let the number of colors be $N = 1/p$.

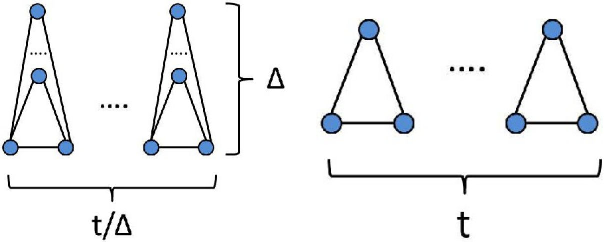
[Pagh and Tsourakakis, 2012]

- Let $f : V \rightarrow [N]$ have uniformly random values
- $E' \leftarrow \{\{u, v\} \in E \mid f(u) = f(v)\}$
- $T \leftarrow$ number of triangles in the graph (V, E')
- **return** T/p^2

Colorful triangle counting

Theorem ([Pagh and Tsourakakis, 2012])

If $p \geq \max\left(\frac{\Delta \log n}{t}, \sqrt{\frac{\log n}{t}}\right)$, then $T \sim \mathbb{E}[T]$ with probability $1 - o(1)$.

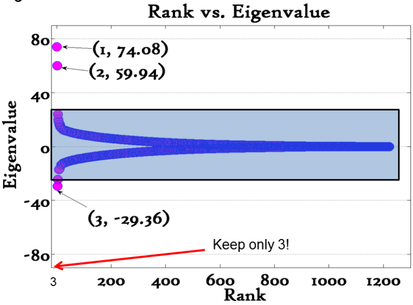


(a) (b)

Figure: Conditions of the Theorem are tight. In order to hope for concentration p has to be greater than (a) $\frac{\Delta}{t}$ and (b) $t^{-1/2}$.

Counting triangles using matrix-vector multiplications [Tsourakakis, 2008]

The number of triangles t_i that vertex i participates in, can be computed from the spectrum of the adjacency matrix $t_i = \frac{\sum_j \lambda_j^3 u_{j,i}^2}{2}$, and the total number of triangles just from the eigenvalues $t = \frac{1}{6} \sum_{i=1}^{n} \lambda_i^3$.



Discussion

- Both [Tsourakakis et al., 2011] and [Pagh and Tsourakakis, 2012] are amenable to distributed implementations (e.g., MapReduce).
- Also, have been adapted to the streaming models.
- A system built which can handle big graph data based on the primitive "matrix-vector" multiplication in MapReduce [Kang and Tsourakakis et al., 2009]

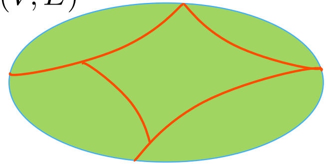
Open Source Software World Challenge, Silver Award and officially included in Hadoop for Windows Azure for Microsoft's big data plans

http://www.cs.cmu.edu/~pegasus/

Balanced graph partitioning

- Graph has to be distributed across a cluster of machines

$G = (V, E)$



- graph partitioning is a way to split the graph vertices in multiple machines
- graph partitioning objectives guarantee low communication overhead among different machines
- additionally balanced partitioning is desirable
- each partition contains $\approx n/k$ vertices, where n, k are the total number of vertices and machines respectively

Off-line k-way graph partitioning

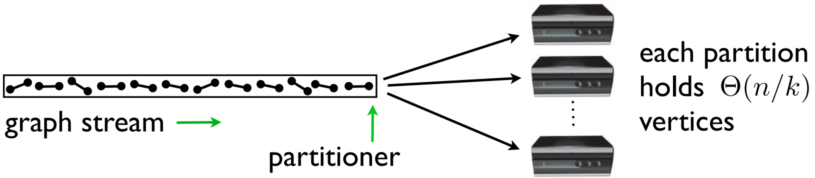
problem: minimize number of edges cut, subject to cluster sizes being at most $\nu n/k$ (bi-criteria approximations)

- $\nu = 2$: Krauthgamer, Naor and Schwartz [Krauthgamer et al., 2009] provide $O(\sqrt{\log k \log n})$ approximation ratio based on the work of Arora-Rao-Vazirani for the sparsest-cut problem ($k = 2$) [Arora et al., 2009]
- $\nu = 1 + \epsilon$: Andreev and Räcke [Andreev and Räcke, 2006] combine recursive partitioning and dynamic programming to obtain $O(\epsilon^{-2} \log^{1.5} n)$ approximation ratio.

Practice: METIS algorithm [Karypis and Kumar, 1998], not well understood but performs well.

Streaming *k*-way graph partitioning

- input is a data stream
- graph is ordered
  - arbitrarily
  - breadth-first search
  - depth-first search
- generate an approximately balanced graph partitioning



Graph representations

- incidence stream
  - at time t, a vertex arrives with its neighbors
- adjacency stream
  - at time t, an edge arrives

Partitioning strategies

- hashing: place a new vertex to a cluster/machine chosen uniformly at random
- neighbors heuristic: place a new vertex to the cluster/machine with the maximum number of neighbors
- non-neighbors heuristic: place a new vertex to the cluster/machine with the minimum number of non-neighbors

FENNEL algorithm [Tsourakakis et al., 2014]

The standard formulation hits the ARV barrier

$$\begin{aligned} \text{minimize}_{\mathcal{P}=(S_1,\ldots,S_k)} \quad & |\partial\, e(\mathcal{P})| \\ \text{subject to} \quad & |S_i| \leq \nu \frac{n}{k}, \text{ for all } 1 \leq i \leq k \end{aligned}$$

- We relax the hard cardinality constraints

$$\text{minimize}_{\mathcal{P}=(S_1,\ldots,S_k)} \quad |\partial\, E(\mathcal{P})| + c_{\text{IN}}(\mathcal{P})$$

where $c_{\text{IN}}(\mathcal{P}) = \sum_i s(|S_i|)$, so that objective self-balances

FENNEL algorithm

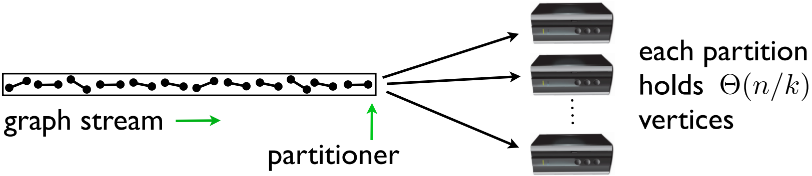
- for $S \subseteq V$, $f(S) = e[S] - \alpha|S|^\gamma$, with $\gamma \geq 1$ (related to optimal quasicliques [Tsourakakis et al., 2013])
- given partition $\mathcal{P} = (S_1, \ldots, S_k)$ of V in k parts define

$$g(\mathcal{P}) = f(S_1) + \ldots + f(S_k)$$

- the goal: maximize $g(\mathcal{P})$ over all possible k-partitions
- notice:

$$g(\mathcal{P}) = \underbrace{\sum_i e[S_i]}_{m-\text{number of edges cut}} - \underbrace{\alpha \sum_i |S_i|^\gamma}_{\text{minimized for balanced partition!}}$$

FENNEL algorithm — greedy scheme



- send v to the partition / machine that maximizes

$$
\begin{aligned}
f(S_i \cup \{v\}) - f(S_i) \\
&= e[S_i \cup \{v\}] - \alpha(|S_i| + 1)^\gamma - (e[S_i] - \alpha|S_i|^\gamma) \\
&= d_{S_i}(v) - \alpha\mathcal{O}(|S_i|^{\gamma-1})
\end{aligned}
$$

- fast, amenable to streaming and distributed setting

FENNEL algorithm

Theorem

- For $\gamma = 2$ there exists an approximation algorithm that achieves an approximation factor $\log(k)/k$
- random partitioning gives approximation factor $1/k$
- no dependence on n

  mainly because of relaxing the hard cardinality constraints
- $\gamma = 2$ gives non-neighbors heuristic
- $\gamma = 1$ gives neighbors heuristic
- interpolate between the two heuristics, e.g., $\gamma = 1.5$
- The algorithm can recover the true partitions under various random graph models in sublinear time [Tsourakakis, 2014] using higher length random walks.

FENNEL algorithm — results

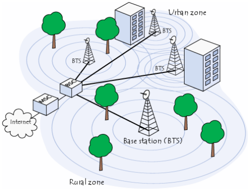
Twitter graph with approximately 1.5 billion edges, $\gamma = 1.5$

$$\lambda = \frac{\#\{\text{edges cut}\}}{m} \qquad \rho = \max_{1 \leq i \leq k} \frac{|S_i|}{n/k}$$

	Fennel		Best competitor		Hash Partition		METIS	
k	λ	ρ	λ	ρ	λ	ρ	λ	ρ
2	6.8%	1.1	34.3%	1.04	50%	1	11.98%	1.02
4	29%	1.1	55.0%	1.07	75%	1	24.39%	1.03
8	48%	1.1	66.4%	1.10	87.5%	1	35.96%	1.03

Table: Fraction of edges cut λ and the normalized maximum load ρ for Fennel, the best competitor and hash partitioning of vertices for the Twitter graph. Fennel and best competitor require around 40 minutes, METIS more than $8\frac{1}{2}$ hours.

Rainbow connection



- Suppose we wish to route messages in a cellular network G, between any two vertices in a pipeline, and require that each link on the route between the vertices (namely, each edge on the path) is assigned a distinct channel (e.g., a distinct frequency). The minimum number of distinct channels we need to use is the rainbow connectivity of G.

Rainbow connection

- An edge colored graph G is rainbow edge connected iff any two vertices are connected by a path whose edges have distinct colors. The rainbow connectivity $rc(G)$ of a connected graph G is the smallest number of colors that are needed in order to make G rainbow edge connected.
- $rc(G) \leq n - 1$
- $rc(G) = n - 1$ iff G is a tree
- $rc(G) = 1$ iff G is the complete graph K_n
- $rc(G) \leq n\frac{4 \log n + 3}{\delta}$ [Caro et al., 2008]

Rainbow connection of sparse random graphs

Let

$$L = \frac{\log n}{\log \log n} \tag{3}$$

and let $A \sim B$ denote $A = (1 + o(1))B$ as $n \to \infty$.

Theorem ([Frieze and Tsourakakis, 2012a, Frieze and Tsourakakis, 2012b])

Let $G = G(n, p), p = \frac{\log n + \omega}{n}, \omega \to \infty, \omega = o(\log n)$. Also, let Z_1 be the number of vertices of degree 1 in G. Then, with high probability(whp)

$$rc(G) \sim \max\{Z_1, L\},$$

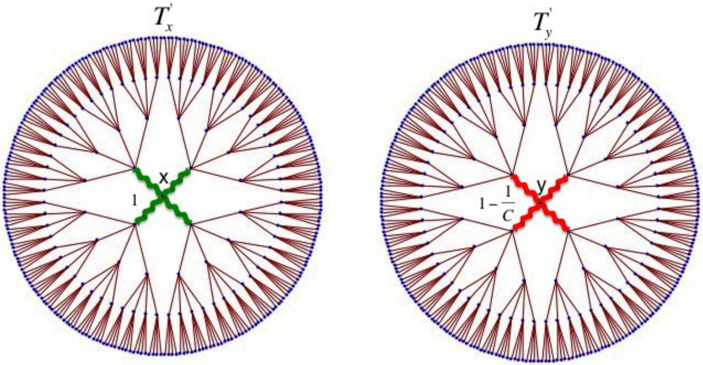
Rainbow connection of sparse random graphs

Let a vertex be *large* if $\deg(x) \geq \log n/100$ and *small* otherwise.

(Very) High-level sketch of the proof

To prove that this works, we have to find, for each pair of *large* vertices x, y, a large collection of edge disjoint paths joining them.

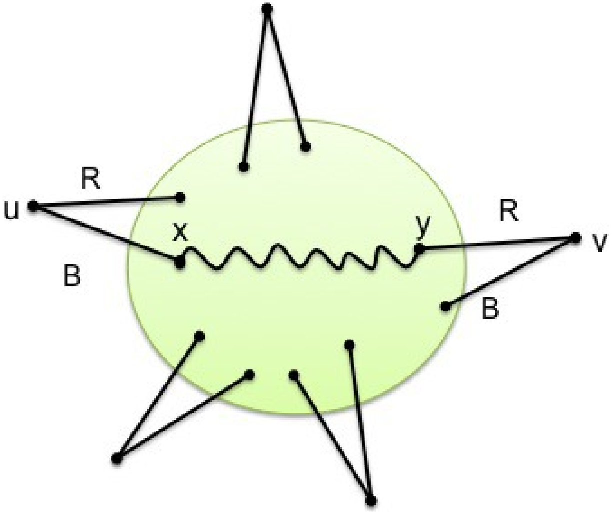
Rainbow connection of sparse random graphs



Top-down coloring, think of it as an evolutionary process. We show that there are many "alive" pairs.

Rainbow connection of sparse random graphs

Taking care of small vertices.



Rainbow connection of sparse random graphs

Also results for random regular graphs. Random coloring does not work now!

Theorem ([Dudek Frieze Tsourakakis, 2013, Frieze and Tsourakakis, 2012b])

Let $G = G(n, r)$ be a random r-regular graph where $r \geq 3$ is a fixed integer. Then, whp

$$rc(G) = \begin{cases} O(\log^4 n) & r = 3 \\ O(\log n) & r \geq 4. \end{cases}$$

Few open problems

- Random Apollonian networks: Conductance $\phi(G)$. We conjecture $\phi(G) = \Theta(\frac{1}{\sqrt{n}})$ **whp** .
- Triangle counting: Same problem in other computational models (I/O efficient).
- Rainbow connectivity: Random 3-regular graphs case remains open. We conjecture $rc(G) = \log n$ in this case as well.

For more related open problems, see Chapter 13 in [Tsourakakis, 2013].

Research directions

- Modeling: How can the Internet and the power grid, two highly engineered networks, be represented by the same graph model of a social network? Narrow down the class of networks of interest!
- Implications: Assume that a good model is available. Can we use the special properties of this model to come up with efficient algorithms for important graph-structured computations? Can the analysis of certain random processes create value out of data?
- Scalability: Algorithms and systems that allow us to large-scale networks.
- Reconstruction problems: A lot of interest in cancer phylogenetics.

Research goal: Develop principled approaches that create value out of data based on well-founded mathematical, statistical and algorithmic tools.

thank you!

references I

Andrade Jr, J. S., Herrmann, H. J., Andrade, R. F., and da Silva, L. R. (2005).
Apollonian networks: Simultaneously scale-free, small world, euclidean, space filling, and with matching graphs.
Physical Review Letters, 94(1):018702.

Andreev, K. and Räcke, H. (2006).
Balanced graph partitioning.
Theor. Comp. Sys., 39(6):929–939.

Arora, S., Rao, S., and Vazirani, U. (2009).
Expander flows, geometric embeddings and graph partitioning.
Journal of the ACM (JACM), 56(2).

Bollobás, B. and Riordan, O. (2003).
Mathematical results on scale-free random graphs.
Handbook of graphs and networks, 1:34.

references II

Buckley, P. G. and Osthus, D. (2004).
Popularity based random graph models leading to a scale-free degree sequence.
Discrete Mathematics, 282(1):53–68.

Caro, Y., Lev, A., Roditty, Y., Tuza, Z., and Yuster, R. (2008).
On rainbow connection.
Electron. J. Combin, 15(1):R57.

Dudek Frieze Tsourakakis, . (2013).
Rainbow connection of random regular graphs.
arXiv preprint arXiv:1311.2299.

Fortunato, S. (2010).
Community detection in graphs.
Physics Reports, 486(3):75–174.

references III

Frieze, A. and Tsourakakis, C. E. (2013).
Some properties of random apollonian networks.
Internet Mathematics.

Frieze, A. M. and Tsourakakis, C. E. (2012a).
Rainbow connection of sparse random graphs.
The Electronic Journal of Combinatorics, 19.

Frieze, A. M. and Tsourakakis, C. E. (2012b).
Rainbow connectivity of sparse random graphs.
In *APPROX-RANDOM*, pages 541–552.

Holme, P. and Kim, B. J. (2002).
Growing scale-free networks with tunable clustering.
Physical Review E, 65(2):026107.

references IV

Kang, U. and Tsourakakis et al., C. (2009).

Pegasus: A peta-scale graph mining system implementation and observations.

In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pages 229–238. IEEE.

Karypis, G. and Kumar, V. (1998).

A fast and high quality multilevel scheme for partitioning irregular graphs.

SIAM J. Sci. Comput., 20(1):359–392.

Krauthgamer, R., Naor, J. S., and Schwartz, R. (2009).

Partitioning graphs into balanced components.

In *SODA*.

references V

Lipton, R. J. and Tarjan, R. E. (1979).
A separator theorem for planar graphs.
SIAM Journal on Applied Mathematics, 36(2):177–189.

Pagh, R. and Tsourakakis, C. E. (2012).
Colorful triangle counting and a mapreduce implementation.
Information Processing Letters, 112(7):277–281.

Saramäki, J. and Kaski, K. (2004).
Scale-free networks generated by random walkers.
Physica A: Statistical Mechanics and its Applications, 341:80–86.

Tsourakakis, C., Kolountzakis, M., and Miller, G. (2011).
Triangle sparsifiers.
Journal of Graph Algorithms and Applications, 15(6).

references VI

Tsourakakis, C. E. (2008).

Fast counting of triangles in large real networks without counting: Algorithms and laws.

In *ICDM*.

Tsourakakis, C. E. (2013).

Mathematical and Algorithmic Analysis of Network and Biological Data.

PhD thesis, Carnegie Mellon University.

Tsourakakis, C. E. (2014).

Egypt: Efficient graph partitioning for large distributed graphs.

references VII

Tsourakakis, C. E., Bonchi, F., Gionis, A., Gullo, F., and Tsiarli, M. A. (2013).

Denser than the densest subgraph: Extracting optimal quasi-cliques with quality guarantees.

KDD.

Tsourakakis, C. E., Gkantsidis, C., Radunovic, B., and Vojnovic, M. (2014).

FENNEL: Streaming graph partitioning for massive scale graphs.

In *WSDM*.