Complexity of Combinatorial Optimization in Power Law Graphs

Gopal Pandurangan

Department of Computer Science
Purdue University

Joint work with
Alessandro Ferrante (U. of Salerno)
Kihong Park (Purdue)

Algorithmic Complexity of Real-World Graphs

- Is combinatorial optimization easy or hard in real-world graphs?
- Real-world graphs possess special properties which may very well make the problems tractable.

Power Law Graphs

Many large-scale real-world graphs exhibit a power law distribution.

The number of nodes y_i of a given degree i is proportional to $i^{-\beta}$ where $\beta > 0$ is a constant.

Power Law in Real-World Graphs

Power-law degree distribution has been observed in:

- Internet — 2.1
- World Wide Web — 2.1
- Social networks: movie actors graph — 2.3; citation graph — 3
- Biological networks: protein-protein interaction graphs — 2.5.

In most real-world graphs, β ranges between 1 and 4.

Power-law graphs have emerged as a partial answer to the perennial search for representative real-world graphs in combinatorial optimization.

Combinatorial Optimization in Real-World Power Law Graphs

Practical evidence suggests that combinatorial optimization in real-world power law graphs is easier than in general graphs.

Internet

Experiments in Internet measurement graphs show that a simple greedy algorithm that exploits the power law property yields a very good approximation to the MINIMUM VERTEX COVER problem. [Park and Lee, 2001]

Social Networks

Experiments show that a simple greedy algorithm that exploits the power law property gives almost optimal solution to the DOMINATING SET problem. [Eubank et al. 2004]

Algorithmic Theory of Power Law Graphs

Above results on disparate problems on various real-world power law graphs motivate a coherent and systematic algorithmic theory of optimization in power law graphs.

Power Law Graph Models

Power law graph models have been proposed to capture/explain the empirically observed power-law degree distribution in real-world graphs.

Can be classified into two types:

1. Takes a power-law degree sequence and generates graph instances with this distribution.

2. Arises from attempts to explain the power-law starting from basic assumptions about a growth evolution.

Aiello, Chung, and Lu (ACL) Model

A robust and general model for (undirected) power-law graphs:

the number of vertices y_i with degree i is (roughly) given by $y_i = e^{\alpha}/i^{\beta}$,

where e^{α} is a normalization constant.

The structural properties that is true in this model will be true for all graphs with the given degree sequence.

We investigate the complexity of simple power-law graphs (no multi-edges or self-loops).

Algorithmic Complexity of Power Law Graphs

Implications of power-law degree distribution to the algorithmic complexity of NP-hard optimization problems.

- Can power-law property alone be sufficient to design polynomial-time algorithms for NP-hard problems on power-law graphs?
- Does the answer depend on the power-law exponent β?

Worst-Case Complexity

All NP-hard graph-theoretic optimization problems that satisfy an optimal substructure property remain NP-Hard on simple power law graphs for all $\beta > 0$.

Examples: MINIMUM VERTEX COVER, MAXIMUM INDEPENDENT SET, and MINIMUM DOMINATING SET.

CLIQUE and COLORING (that do not satisfy the optimal substructure property) are also NP-hard on simple power law graphs for all $\beta \geq 1$.

Average-Case Complexity

Problems such as VERTEX COVER, DOMINATING SET, can be solved asymptotically optimally in dense random power law graphs.

For random power law graphs with $\beta < 2$, there exists a polynomial time algorithm that with probability at least $1 - o(1)$ outputs within $o(1)$ of the optimal.

Talk Outline

1. Hardness Results.
2. Average-Case Results.
3. Conclusion.

Optimal Substructure Property

Optimal solution for a problem on given graph is the union of the optimal (sub-)solutions on its maximal connected components.

Let P be an optimization problem which takes a graph as input.

For every input G: every optimum solution of P on G should contain an optimum solution of P on each of G's maximal connected components.

MINIMUM VERTEX COVER satisfies this property.

MINIMUM COLORING and MAXIMUM CLIQUE do not satisfy the property.

Preliminaries: y and d-degree Sequence

Given a graph, we will refer to two types of sequence of integers:

- Y^G — *y*-degree sequence of G: lists the number of vertices of G with a certain degree i.e., the degree distribution.

- D^G — *d*-degree sequence of G: lists the degrees of the vertices of G in non-increasing order i.e., the degree sequence of the graph in non-increasing order.

We use expansion of Y, $EXP(Y^G) = D^G$.

β-Graphs

The ACL model of power-law graphs have a particular kind of y-degree sequence which we call as (β, α)-degree sequence:

Definition 1. [(β, α)-degree sequence] *Given $\alpha, \beta \in \mathbb{R}^+$, the y-degree sequence of a graph $G = (V, E)$ is a (β, α)-degree sequence denoted by $Y^{(\beta,\alpha)} =< y_1^{(\beta,\alpha)}, \cdots, y_m^{(\beta,\alpha)} >$, if $m = \lfloor e^{\alpha/\beta} \rfloor$ and, for $i \in [1, m]$*

$$y_i = \begin{cases} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor & \text{if } i > 1 \text{ or } \sum_{k=1}^{m} \left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor \text{ is even} \\ \lfloor e^\alpha \rfloor + 1 & \text{otherwise.} \end{cases}$$

Definition 2. [β-graph] *Given $\beta \in \mathbb{R}^+$, a graph $G = (V, E)$ is a β-graph if it is simple and there exists $\alpha \in \mathbb{R}^+$ such that the y-degree sequence of G is a (β, α)-degree sequence.*

Embedding Lemma

A key technical lemma that is needed for hardness results.

Shows the existence of a β-graph with a certain property.

Lemma 1. [Embedding Lemma] *Let* $G = (V, E)$ *be a simple undirected graph and* $\beta \in \mathbb{R}^+$*. Then there exists a simple undirected graph* $G_1 = (V_1, E_1)$ *such that* G *is a set of maximal connected components of* G_1*,* $|V_1| = poly(|V|)$ *and* G_1 *is a* β*-graph. Furthermore, given* G*, we can construct* G_1 *in time polynomial in the size of* G*.*

The difficulty is in proving that such a degree sequence exists.

NP-Hardness

Theorem 1. *Let P be an optimization problem on graphs with the optimal substructure property. If P is NP-hard on (simple) general graphs, then it is also NP-hard on β-graphs for all $\beta > 0$.*

Erdos-Gallai Condition

A degree sequence is called graphic if it can be realized as simple graph (no self-loops or multi-edges).

Definition 3. [Eligible Sequence] *A sequence of integers* $S =< s_1, \cdots, s_n >$ *is eligible if* $s_1 \geq \cdots \geq s_n$ *and, for all* $k \in [1, n]$, $f_S(k) \geq 0$, *where*

$$f_S(k) = k(k-1) + \sum_{i=k+1}^{n} \min\{k, s_i\} - \sum_{i=1}^{k} s_i.$$

Lemma 2. [Erdos and Gallai] *A sequence of integers* S *is graphic if and only if it is non-increasing,* $tot(S) = \sum_{i=1}^{n} s_i$ *is even and* S *is eligible.*

The weaker condition $\max_i d_i^2 < \sum_i d_i$ and $\sum_i d_i$ is even only ensures that the degree sequence can be realized as a multi-graph.

Havel and Hakimi Condition

Lemma 3. [Havel and Hakimi] *A sequence of integers* $D =< d_1, \cdots, d_n >$ *is graphic if and only if it is non-increasing, and the sequence of values* $D' =< d_2 - 1, d_3 - 1, \cdots, d_{d_1+1} - 1, d_{d_1+2}, \cdots, d_n >$ *when sorted in non-increasing order is graphic.*

Gives a straightforward algorithm to construct a simple graph from a graphic degree sequence.

Embedding Lemma

Given two degree sequences $S =< s_1, \cdots, s_n >$ and $T =< t_1, \cdots, t_m >$ with $n \geq m$, we will denote $S - T =< x_1, \cdots, x_n >$ with $x_i = s_i - t_i$ if $i \in [1, m]$ and $x_i = s_i$ otherwise.

To prove the existence of the β-graph G_1:

Given $\beta \in \mathbb{R}^+$, there exists $\alpha_0 \in \mathbb{R}^+$ such that, for all $\alpha \geq \alpha_0$, the sequence $Y = Y^{(\beta,\alpha)} - Y^G$ is the y-degree sequence of a simple graph.

To show this, we will show that the expansion of the prefix $< y_1, \cdots, y_m >$ of Y is eligible for $\alpha \geq \alpha_1$, and that the expansion of the suffix $< y_{m+1}, \cdots, y_{m_1} >$ is eligible for $\alpha \geq \alpha_2$, where m is the maximum degree of G and m_1 is the number of elements in $Y^{(\beta,\alpha)}$.

These give a sufficient condition for $EXP(Y)$ to be graphic.

CLIQUE and COLORING

- Given an arbitrary graph G, construct a simple β-graph G_1 containing G as a maximal connected component.

  Let $G_2 = G_1 \backslash G$ be the remaining graph.

- Show that problem can be solved optimally in G_2.

- Construct G_1 (and hence G_2) in such a way that it admits an efficient algorithm.

- Make G_2 to be a simple bipartite graph.

Embedding Lemma for $\beta \geq 1$

Lemma 4. *Let $G = (V, E)$ be a simple graph. Then for all $\beta >= 1$, there exists a β-graph G_1 such that $G_2 = G_1 - G$ is a bipartite graph. The size of G_1 is at most polynomial in the size of G.*

Corollary 1. *CLIQUE, and COLORING are NP-Complete in β-graphs for all $\beta \geq 1$.*

Bipartite Graph Construction

A degree sequence $D =< d_1, \cdots, d_n >$ with maximum value m is contiguous if $y_i > 0$ for all $i \in [1, m]$.

Lemma 5. *Let $D =< d_1, \cdots, d_n >$ be a degree sequence. If D is contiguous and $d_1 \leq \lfloor n/2 \rfloor$ then there exist a simple bipartite graph $G = (V, E)$ such that $D^G = D$. (Bipartite-eligible)*

Proof by Algorithm

Algorithm outputs the graph $G = (S \cup T, E)$:

1. Assign vertices to S and T:

   Starting from the highest vertex, place vertices successively in S and T whichever set has the lower total degree currently.

2. Assign edges between S and T:

   Residual degree of a vertex: final degree - "current" degree.

   Initially all the vertices have degree 0.

3. **repeat** till possible:

   3.1 Assign all edges between the top (highest residual degree) vertex of S and the top vertices of T.
   3.2 Assign all edges between the top vertex of T and the top vertices of S.

Correctness of Algorithm

The residual degree of the set S (T respectively) at the beginning of round i is denoted by $R_i(S)$ ($R_i(T)$ respectively). The number of vertices with positive residual degree (*non full* vertices) in S (T) is denoted by $N_i(S)$ ($N_i(T)$).

The residual degrees of set S at round i in decreasing order is $s_1^i, \cdots, s_h^i$ and those of set T is $t_1^i, \cdots, t_k^i$.

The proof is by induction on the round i. The following invariant holds:

1. $R_i(S) = R_i(T)$ and
2. $N_i(T) \geq d(s_1^i)$ and $N_i(S) \geq d(t_1^i)$.

Choosing a suitable β-graph

Lemma 6. *Let* $G = (V, E)$ *be a simple graph with* n_1 *vertices and* $\beta \geq 1$*. Let* $\alpha_0 = \max\{4\beta, \beta \ln n_1 + \ln(n_1 + 1)\}$*. Then, for all* $\alpha \geq \alpha_0$ *the sequence* $D = EXP(Y^{(\beta,\alpha)} - Y^G))$ *is bipartite-eligible.*

Proof

Let n_2 be the number of elements in D and $\alpha \geq \alpha_0$. We have

$$n_2 \geq \sum_{i=1}^{\lfloor e^{\alpha/\beta} \rfloor} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor - n_1 > e^\alpha \sum_{i=1}^{\lfloor e^{\alpha/\beta} \rfloor} \frac{1}{i^\beta} - \lfloor e^{\alpha/\beta} \rfloor - n_1$$

$$\geq e^\alpha \int_{i=1}^{\lfloor e^{\alpha/\beta} \rfloor + 1} \frac{1}{i^\beta} - \lfloor e^{\alpha/\beta} \rfloor - n_1.$$

If $\beta = 1$, then we have

$$n_2 \geq \alpha e^\alpha - e^\alpha - n_1 \geq 4e^\alpha - 2e^\alpha + 1 \geq 2m + 1$$

and if $\beta > 1$ we have

$$n_2 \geq \frac{e^\alpha}{\beta - 1} - e^{\alpha/\beta} - n_1 \geq 4e^{\alpha/\beta} - 2e^{\alpha/\beta} + 1 \geq 2m + 1,$$

$$y_{n_1}^{(\beta,\alpha)} \geq \left\lfloor \frac{e^\alpha}{n_1^\beta} \right\rfloor > \frac{e^\alpha}{n_1^\beta} - 1 \geq \frac{n_1^{\beta+1} + n_1^\beta}{n_1^\beta} - 1 = n_1,$$

A Greedy Algorithm for VERTEX-COVER

1. Choose the top degree vertices that covers at least $(1 - \delta(n))$ fraction of the total number of edges for an appropriately chosen function $\delta(n) = o(1)$.

2. Include vertices that covers the remaining uncovered edges.

Average-Case Analysis

A random power law graph model: A random graph model that generates graphs whose expected degree-sequence follows a given power law degree sequence. [Lu and Chung, 2002]

Given a power-law degree sequence $d_1, \ldots, d_n$, an edge occurs between vertices i and j with probability

$$\frac{d_i d_j}{\sum_{k=1}^{n} d_k}.$$

A generalization of the classical Erdos-Renyi model.

Theorem 2. *For a random power law graph with $\beta \leq 2$, there exists a polynomial time algorithm that with probability at least $1 - o(1)$ outputs a $1 + o(1)$ approximation to the VERTEX COVER problem.*

Conclusion

- Real-world power law graphs are not "worst-case" instances of power law graphs, but rather typical instances, and may be well-modeled by power law random graph models.

- A general technique for establishing NP-hardness of a large class of problems in power-law graphs.

- Embedding Lemma can have other applications, e.g., in showing hardness of approximation in power-law graphs

- On the positive side, investigate approximation algorithms for sparse power law graphs.