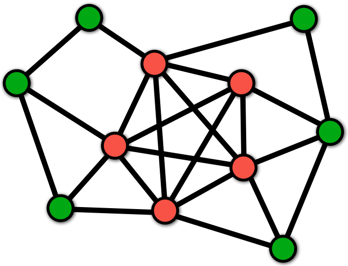
Densest subgraph sparsifiers

densest subgraph problem

- undirected graph $G = (V, E)$
- subgraph induced by $S \subseteq V$
- clique: Let all red nodes be set S (they form a clique). The average degree of S is 4, as each node is connected to four other nodes in S. Also, we can see that by $\frac{2\times\binom{5}{2}}{5} = 4$



density measures[1]

- edge density (average degree):

$$d(S) = \frac{2\,|E(S,S)|}{|S|} = \frac{2\,|E(S)|}{|S|}$$

(sometimes just drop 2)

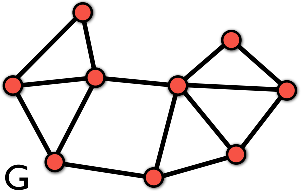
- degree density (half of average degree):

$$\rho(S) = \frac{|E(S)|}{|S|}$$

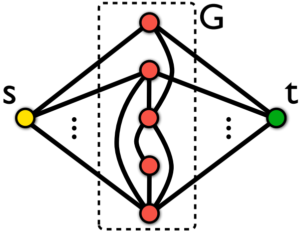
[1]More in DSD module

Goldberg's algorithm for densest subgraph

- consider first edge density d



- is there a subgraph S with $d(S) \geq c$?
- transform to a min-cut instance



- on the transformed instance:
- is there a cut smaller than a certain value?

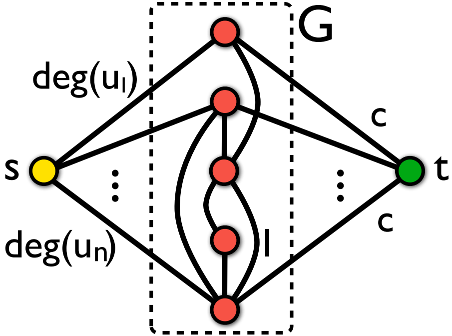
Goldberg's algorithm for densest subgraph

is there S with $d(S) \geq c$?

$$
\begin{aligned}
\frac{2\,|E(S,S)|}{|S|} &\geq c \\
2\,|E(S,S)| &\geq c|S| \\
\sum_{u\in S} \deg(u) - |E(S,\bar{S})| &\geq c|S| \\
\sum_{u\in S} \deg(u) + \sum_{u\in \bar{S}} \deg(u) - \sum_{u\in \bar{S}} \deg(u) - |E(S,\bar{S})| &\geq c|S| \\
\sum_{u\in \bar{S}} \deg(u) + |E(S,\bar{S})| + c|S| &\leq 2\,|E|
\end{aligned}
$$

Goldberg's algorithm for densest subgraph

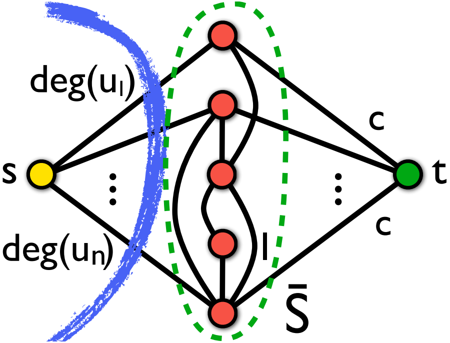
- transformation to min-cut instance



- is there S s.t. $\sum_{u \in \bar{S}} \deg(u) + |e(S, \bar{S})| + c|S| \leq 2\,|E|$?

Goldberg's algorithm for densest subgraph

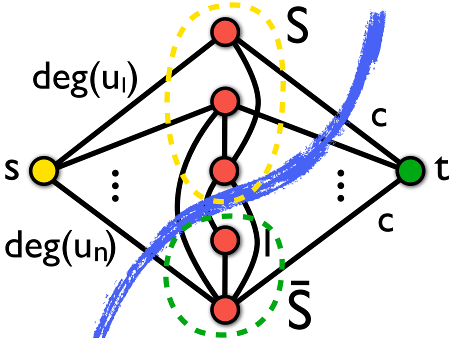
- transform to a min-cut instance



- is there S s.t. $\sum_{u \in \bar{S}} \deg(u) + |e(S, \bar{S})| + c|S| \leq 2\,|E|$?
- a cut of value $2\,|E|$ always exists, for $S = \emptyset$

Goldberg's algorithm for densest subgraph

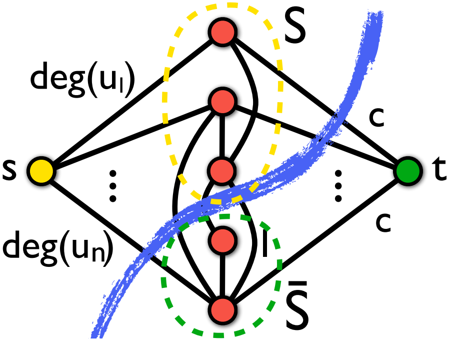
- transform to a min-cut instance



- is there S s.t. $\sum_{u \in \bar{S}} \deg(u) + |e(S, \bar{S})| + c|S| \leq 2\,|E|$?
- $S \neq \emptyset$ gives cut of value $\sum_{u \in \bar{S}} \deg(u) + |e(S, \bar{S})| + c|S|$

Goldberg's algorithm for densest subgraph

- transform to a min-cut instance



- is there S s.t. $\sum_{u \in \bar{S}} \deg(u) + |e(S, \bar{S})| + c|S| \leq 2\,|E|$?
- YES, if min cut achieved for $S \neq \emptyset$

Goldberg's algorithm for densest subgraph

[Goldberg, 1984]

input: undirected graph $G = (V, E)$, number c
output: S, if $d(S) \geq c$

transform G into min-cut instance $G' = (V \cup \{s\} \cup \{t\}, E', w')$
find min cut $\{s\} \cup S$ on G'
if $S \neq \emptyset$ return S
else return NO

- to find the densest subgraph perform binary search on c
- logarithmic number of min-cut calls
- problem can also be solved with one min-cut call using the parametric max-flow algorithm

densest subgraph problem – discussion

- Goldberg's algorithm polynomial algorithm, but
- $\mathcal{O}(nm)$ time for one min-cut computation
- not scalable for large graphs (millions of vertices / edges)

- We will see more algorithms and formulations for dense subgraph discover ylater in class.

Bibliographic remark

Three different papers roughly at the same time came up with the densest subgraph sparsifier theorem:

- Scalable Large Near-Clique Detection in Large-Scale Networks via Sampling by Mitzemacher et al. **KDD 2015**
- Densest subgraph in dynamic graph streams by McGregor et al. **MFCS 2015**
- Applications of Uniform Sampling: Densest Subgraph and Beyond by Esfandiari et al. **SPAA 2016**

Densest subgraph sparsification

Theorem (Mitzenmacher-Pachocki-Peng-Tsourakakis-Xu)

Let $\epsilon > 0$ be an accuracy parameter. Suppose we sample each edge $e \in E_{\mathcal{H}}$ independently with probability $p_D = C\frac{\log n}{D}$ where $D \geq \log n$ is the density threshold parameter and $C = \frac{6}{\epsilon^2}$ is a constant depending on ϵ. Then, the following statements hold simultaneously with high probability: (i) For all $U \subseteq V$ such that $\rho(U) \geq D$, $\tilde{\rho}(U) \geq (1-\epsilon)C \log n$ for any $\epsilon > 0$. (ii) For all $U \subseteq V$ such that $\rho(U) < (1-2\epsilon)D$, $\tilde{\rho}(U) < (1-\epsilon)C \log n$ for any $\epsilon > 0$.

Graph Spanners

- We will follow Stefano Leucci's notes, available at
  `https://www.mpi-inf.mpg.de/fileadmin/inf/d1/teaching/summer19/algorithms/lecture_notes_spanners.pdf`

- Suggested optional readings
  1. [Abboud and Bodwin, 2017]
  2. [Aingworth et al., 1999]
  3. [Althöfer et al., 1993]
  4. [Baswana and Sen, 2007]
  5. [Peleg and Schäffer, 1989]

references I

Abboud, A. and Bodwin, G. (2017).
The 4/3 additive spanner exponent is tight.
Journal of the ACM (JACM), 64(4):1–20.

Aingworth, D., Chekuri, C., Indyk, P., and Motwani, R. (1999).
Fast estimation of diameter and shortest paths (without matrix multiplication).
SIAM Journal on Computing, 28(4):1167–1181.

Althöfer, I., Das, G., Dobkin, D., Joseph, D., and Soares, J. (1993).
On sparse spanners of weighted graphs.
Discrete & Computational Geometry, 9(1):81–100.

references II

Angel, A., Sarkas, N., Koudas, N., and Srivastava, D. (2012).
Dense subgraph maintenance under streaming edge weight updates for real-time story identification.
Proceedings of the VLDB Endowment, 5(6):574–585.

Baswana, S. and Sen, S. (2007).
A simple and linear time randomized algorithm for computing sparse spanners in weighted graphs.
Random Structures & Algorithms, 30(4):532–563.

Beutel, A., Xu, W., Guruswami, V., Palow, C., and Faloutsos, C. (2013).
Copycatch: stopping group attacks by spotting lockstep behavior in social networks.
In *Proceedings of the 22nd international conference on World Wide Web*, pages 119–130.

references III

Goldberg, A. V. (1984).
Finding a maximum density subgraph.
Technical report, University of California at Berkeley.

Peleg, D. and Schäffer, A. A. (1989).
Graph spanners.
Journal of graph theory, 13(1):99–116.

Sarıyüce, A. E., Seshadhri, C., Pinar, A., and Catalyurek, U. V. (2015).
Finding the hierarchy of dense subgraphs using nucleus decompositions.
In *Proceedings of the 24th International Conference on World Wide Web*, pages 927–937.