COMP229: Introduction to Data Science
Lecture 8: connectivity of graphs

Vitaliy Kurlin, vitaliy.kurlin@liverpool.ac.uk
Autumn 2018, Computer Science department
University of Liverpool, United Kingdom

Statistical models of graphs

To analyse real networks, statistical models of random (non-deterministic) graphs are studied.

Definition 8.1. The *Erdős-Rényi* model: for fixed numbers of vertices and edges, all edges are equally likely (have the same probability).

The *Gilbert* model (more popular now): any edge has a fixed probability $P \in [0, 1]$ to be present in a graph (independently of other edges).

What is the expected number of edges in the Gilbert model for a given probability P?

Maximum number of edges

Claim 8.2. A graph without loops and multiple edges on n vertices has at most $\dfrac{n(n-1)}{2}$ edges.

If a problem sounds hard, simplify, e.g. try small values of parameters: $n = 2$, $n = 3$ etc.

Proof. Any of n vertices has at most $n - 1$ edges: one per each of the remaining $n - 1$ vertices.

Choices of vertices are independent, we multiply: $n \times (n - 1)$ pairs. Divide by 2 as we counted each edge twice: from a vertex A to B, from B to A. $\square$

Expected number of edges

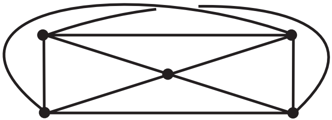
Building a random graph in the Gilbert model is similar to tossing a coin for each of $\dfrac{n(n-1)}{2}$ possible edges. An edge is included into a random graph with a probability $P \in [0, 1]$, this edge is excluded with the probability $1 - P$.

The expected number of edges between 2 vertices is $0 \times (1 - P) + 1 \times P = P$. Taking the sum over $\dfrac{n(n-1)}{2}$ independent outcomes, the *expected* number of edges in the final graph is $\dfrac{n(n-1)}{2}P$.

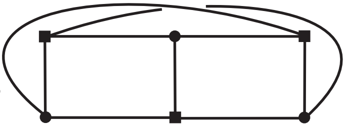
Complete and bipartite graphs

Definition 8.3. A *complete* graph K_n on n vertices has one edge for each (unordered) pair of vertices.

K_5



$K_{3,3}$



K_n has $\dfrac{n(n-1)}{2}$ edges. A *bipartite* graph $K_{m,n}$ has $m+n$ vertices, each of the first m vertices is connected by a single edge to each of the last n vertices. How many edges does $K_{m,n}$ have?

Paths and components

Definition 8.4. A *path* in a graph is a sequence of edges, where any two successive edges are *adjacent* (share a common vertex). Vertices are *path-connected* if they are connected by a path.

Claim 8.5. Path-connectedness is an equivalence relation on all vertices of a graph G whose classes define (connected) *components* of G, which are preserved under any topological equivalence.

How many edges are needed (always necessary) to form a connected graph on n vertices?

Components are well-defined

Proof of Claim 8.5. Check conditions of Def 4.1.

Reflexivity: any vertex is connected to itself.

Symmetry: a path from a vertex A to another vertex B can be reversed to a path from B to A.

Transitivity: if a vertex A is connected to a vertex B by a path, B is connected to C by another path, the union of these paths connects A to C.

Any component is preserved under mergers and subdivisions of edges. Hence the number of components is a topological invariant. □

Necessity for connectivity

Claim 8.6. Any connected graph on n vertices should have at least $n-1$ edges.

Proof that $n-1$ edges are necessary. Starting from n isolated vertices, any new edge can decrease the number of connected components by at most 1 (when two components merge into one).

Since we should go from n components (initial vertices) to 1, we need at least $n-1$ edges. $\square$

$n-1$ edges are necessary, may not be sufficient.

Sufficiency for connectivity

3 edges $(1,2)$, $(2,3)$, $(3,1)$ define a disconnected graph on 4 vertices. Adding the n-th isolated vertex to K_{n-1} gives a disconnected graph.

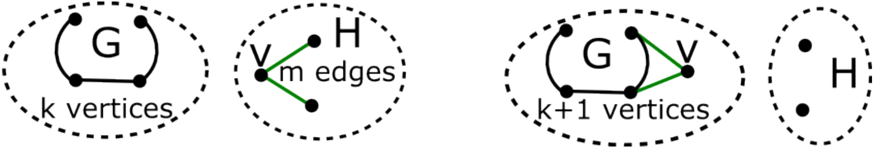
Hence $\dfrac{(n-1)(n-2)}{2}$ edges are not sufficient.

Claim 8.7. Any disconnected graph on n vertices has at most $\dfrac{(n-1)(n-2)}{2}$ edges.

Proof. Without loss of generality we may assume a graph has two connected components: larger G on $k \geq n/2$ vertices, smaller H on $n-k$ vertices.

A proof of sufficiency

A vertex $v \in H$ has $m \leq n-k-1$ edges in H.



Cut the m edges, connect v to $m \leq n-k-1 \leq k$ distinct vertices in G (possible since $2k+1 \geq n$).

The graph is still disconnected with the same number of edges, the smaller component H is even smaller. Continue until H is 1 vertex. Then $G \cup H$ has at most $\dfrac{(n-1)(n-2)}{2}$ edges from K_{n-1}. $\square$

Your questions and the quiz

To benefit from the lecture, now you could

- ask or submit your anonymous questions to the COMP229 folder after the lecture;
- write down your summary in 2–3 phrases, e.g. list key concepts you have learned;
- talk to your classmates to revise the lecture.

Question. How many edges are sufficient to guarantee connectivity of any graph on n vertices?

Answer to the quiz and summary

Answer. By Claim 8.7 the minimum sufficient number of edges to guarantee connectivity of any graph on n vertices is $1 + \dfrac{(n-1)(n-2)}{2}$.

- A graph is called *connected* if any two vertices can be connected by a path edges.
- The number of connected components is a topological invariant of graphs.
- For connectivity of a graph on n vertices, $n - 1$ edges are necessary (not sufficient), $1 + \dfrac{(n-1)(n-2)}{2}$ edges are sufficient.