Random Graphs

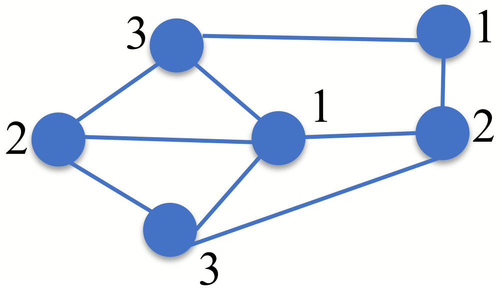
CptS 591: Elements of Network Science

Outline

- Random graph as a concept
- Random variables and Expectation
- Graph invariants in random graphs
- Phase transition
- Random graphs vs real-world networks (subject of next lecture)

Motivation
(from graph theoretic perspective)

- Given a graph G,
  - the minimum length of a cycle contained in G is the *girth* g(G) of G
  - the maximum length of a cycle in G is its *circumference*
  - the smallest number of colors required to color G is the *chromatic number x(G) of G*
- Example:



- g(G) = 3
- circumference(G) = 6
- x(G) = 3

(in general NP-hard to compute)

Motivation (cont'd): Paraphrased Erdos theorem

- There exist graphs whose
  1. girth is arbitrarily large, and
  2. chromatic number is arbitrarily large

- These requirements work against each other:
  A graph with a large girth is tree-like (acyclic) and hence is expected to have small chromatic number

- ➔ a *constructive proof* for Erdos theorem is difficult (if not impossible) to come by

- Instead, Erdos used *Random Graphs* and the "*Probabilistic Method*" to prove such *existence theorems*

What is a random graph?

- Let V be a fixed set of n elements, say $V = \{1,2,\ldots,n\}$.

  Let $\hat{G}$ be the set of all possible graphs on V.

  (Note: there are 2^N possible graphs $N = (n:2)$, where $(n:k)$ denotes n choose k)

- We would like to turn $\hat{G}$ into a probability space and be able to answer such questions as
  - What is the probability that a graph G in $\hat{G}$ has a certain property?
  - What is the expected value of a given invariant on G?

- Consider the following random process of generating G.
  - Let $[V]^2$ denote the set of all pairs of elements drawn from V

    (There are $(n:2)$ possible pairs)
  - For each e in $[V]^2$ decide using a random expt whether or not e shall be an edge of G
  - Perform the expts independently, each time accepting e to be an edge with a fixed probability p, $0 \le p \le 1$

- Now let G_0 be some fixed graph on V with m edges.
- Then,

$$P[G=G_0] = p^m q^{(N-m)} \text{, where } q = 1-p \text{ and } N = (n:2)$$

What is random graph (cont'd)

- One can continue in this way to determine probabilities of all possible elementary events (all m)

➔the probability measure of the desired space $\hat{G}$ is determined

- One can formally show (as in Diestel) that a probability measure on $\hat{G}$ where all individual edges occur *independently* with *probability p* exists.

- With these two assumptions, we can now calculate probabilities in the space $\hat{G} = \hat{G}(n,p)$.

Examples

- Let G in $\hat{G}$, and H be a fixed graph on a subset U of V. Let the number of vertices in H be k, and the number of edges be l.

- Q1: What is P[H is a subgraph of G]?
- Soln 1: Each edge of H occurs independently with a probability of p. Hence the required probability is p^l.

- Q2: What is P[H is an induced subgraph of G]?
- Soln2: This time, in addition to that in Q1, the $r = (k{:}2) – l$ edges missing from H are required to be missing from G too, independently with probability $q = 1-p$.

  Hence the required probability is $p^l q^r$

More interesting examples

- First we define a few notions
  - Independent Set: a set of pairwise non-adjacent vertices
  - Clique: a set of pairwise adjacent vertices
  - The size of the largest IS in a graph G is its independence number $\alpha(G)$
  - The size of the largest clique in a graph is its clique number $\omega(G)$

- Lemma 1: For all integers n, k with $n \geq k \geq 2$, the probability that G in $\hat{G}(n,p)$ has an IS of size k is at most
  - $P[\alpha(G) \geq k] \leq (n:k)q^{(k:2)}$

- Lemma 2:For all integers n, k with $n \geq k \geq 2$, the probability that G in $\hat{G}(n,p)$ contains a clique of size k is at most
  - $P[w(G) \geq k] \leq (n:k)p^{(k:2)}$

Random variables and Expectation

- Let X be a random variable. Let the possible values X can assume be x_1, x_2, …, x_n.
- The expected (or mean) value of X is then
  $E(X) = \Sigma^n_{i=1} P[X = x_i] \bullet x_i$
- Example: die tossing.
  (Let X be a toss. Convince yourself that $E(X) = 3.5$)
- The operator E, expectation, is linear
  - $E(X + Y) = E(X) + E(Y)$ and
  - $E(aX) = a\, E(X)$

  For any random variables X, Y and real number a
- In the context of random graphs, a graph invariant may be interpreted as a nonnegative random variable on $\hat{G}(n,p)$, i.e., as a function $X:\hat{G}(n,p) \rightarrow [0,\infty]$

Graph invariants in random graphs

- The expected value of X is then
  - $E(x) = \Sigma P(\{G\}) X(G)$ (sum over all G in $\hat{G}$)
- Computing the mean of a random variable X can be an effective way to compute the existence of a graph G s.t.
  - (i) $X(G) < a$ for some fixed $a > 0$, and
  - (ii) G has some desired property
- Idea: if $E(X)$ is small, $X(G)$ is small for many of the graphs in $\hat{G}(n,p)$, since $X(G) \geq 0$ for all G in $\hat{G}$. It is then reasonable to expect to find a graph with the desired property among these.
- This idea lies at the heart of many non-constructive existence proofs using random graphs.

Markov's Inequality

- Lemma 3: Let $X \geq 0$ be a random variable on $\hat{G}(n,p)$ and $a > 0$. Then,

  $P[X \geq a] \leq E(X)/a$

- Proof:

  $E(X) = \Sigma\, P(\{G\})\, X(G)$ (sum over G in $\hat{G}(n,p)$)

  $\geq \Sigma\, P(\{G\})\, X(G)$ (sum over G in $\hat{G}(n,p)$ s.t. $X(G) \geq a$)

  $\geq \Sigma\, P(\{G\})\, a$ (since $X(G) \geq a$)

  $= P[X \geq a]\, a$ (since a is constant)

  Rewriting,

  $P[X \geq a] \leq E(X)/a$

The Probabilistic Method

- Basic Idea:
  - To prove the existence of an object with some desired property, define a probability space on some larger class of objects, and then show that an element of this larger space has the desired property
- Illustrate using proof of Erdos’s theorem

Properties of almost all graphs and Phase transition

- Many results concerning "almost all graphs" have the common feature that the value of p (in the space Ĝ(n,p)) plays no role.
- How could this happen?

- Then, what happens if p is allowed to vary with n?

Phase transition

G in Ĝ (n,p)

p =p(n)

Phase transition

G in Ĝ (n,p)

G a.s. has no edges

n^{-2}

p =p(n)

Phase transition

G in Ĝ (n,p)

G a.s. has no edges

n^{-2}

G acquires more and more edges

$n^{-3/2}$

Every component in G a.s. has at least two vertices

p =p(n)

Phase transition

G in Ĝ (n,p)

G a.s. has no edges

n^{-2}

G acquires more and more edges

$n^{-3/2}$

Every component in G a.s. has at least two vertices

n^{-1}

- First cycles are born
- One component outgrows all others

$(\log n)\, n^{-1}$

A “giant component” is born (G is connected)

$p = p(n)$

Phase transition

G in Ĝ (n,p)

G a.s. has no edges

n^{-2}

G acquires more and more edges

$n^{-3/2}$

Every component in G a.s. has at least two vertices

n^{-1}

- First cycles are born
- One component outgrows all others

$(\log n)\ n^{-1}$

A "giant component" is born (G is connected)

$(1+e)\ (\log n)\ n^{-1}$

G a.s. has a Hamilton cycle

$p = p(n)$

Random graphs vs real-world networks

- Degree Distribution
- Average Path Length
- Clustering Coefficient

We will look at these in next lecture