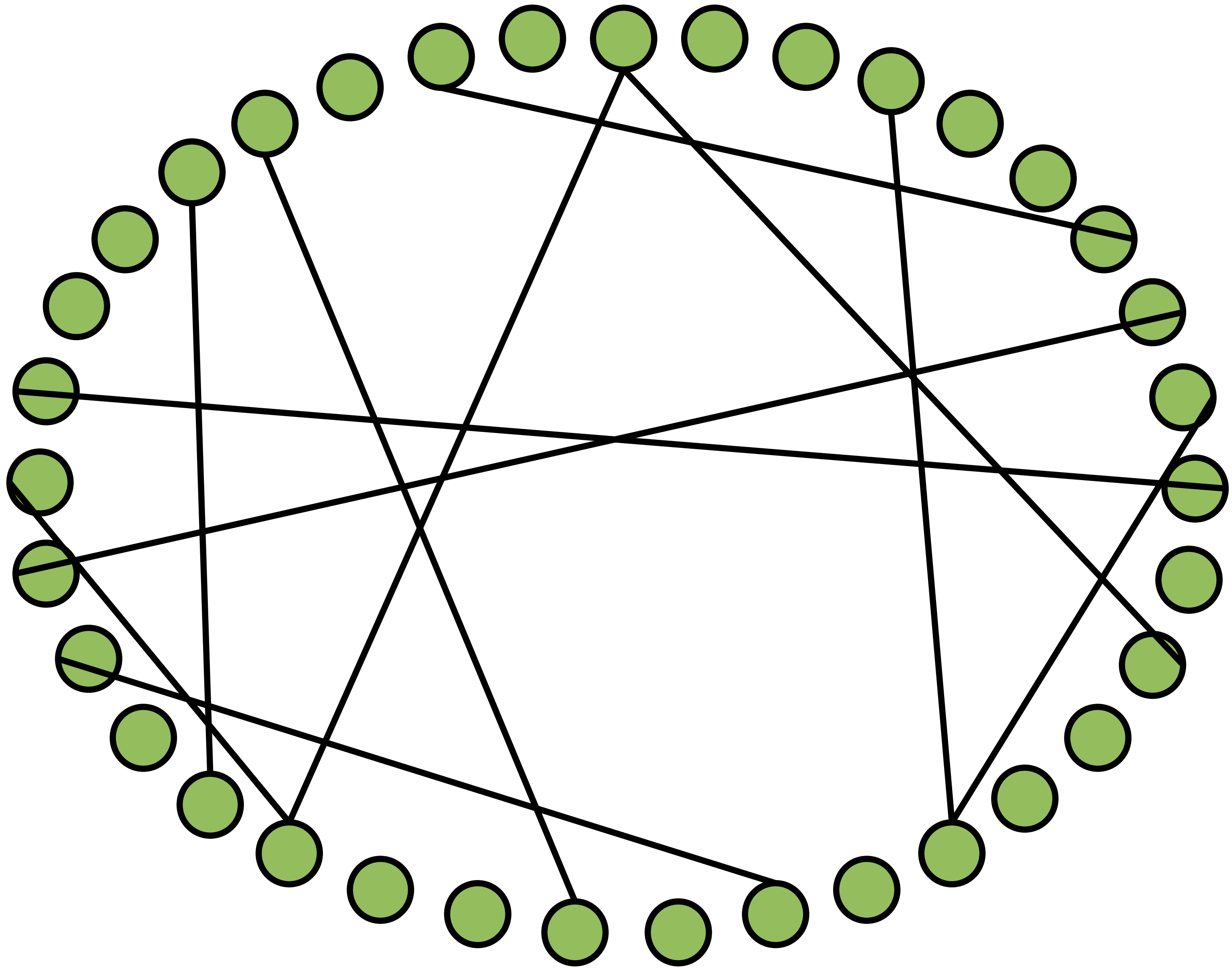
Random Graph G(n,p)



Threshold Functions for Trees and Cycles

- Trees of order k

$$t(n) = n^{\frac{-k}{k-1}}$$

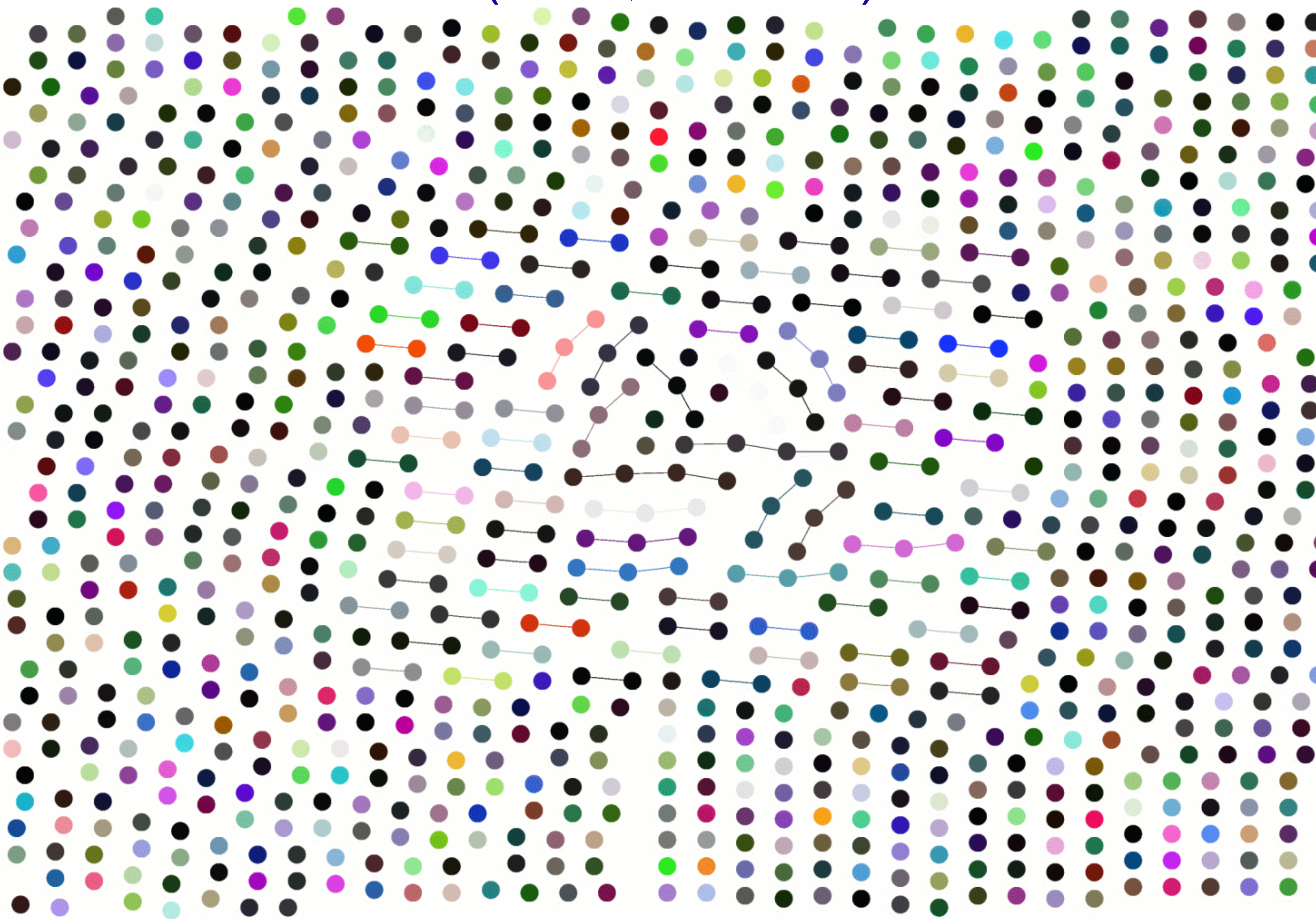
- Cycles of order k

$$t(n) = n^{-1}$$

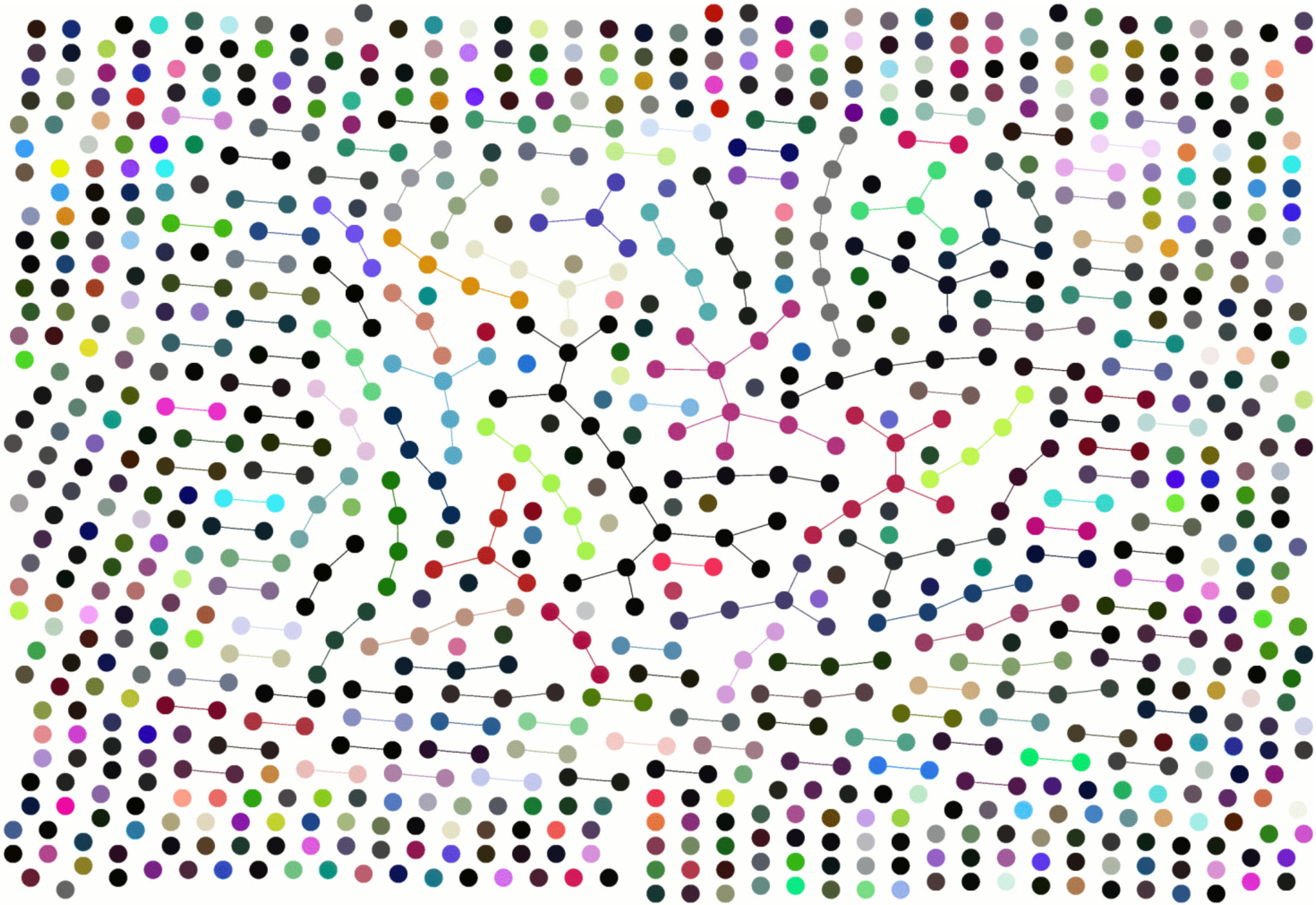
- Complete graphs K_k

$$t(n) = n^{-2/(k-1)}$$

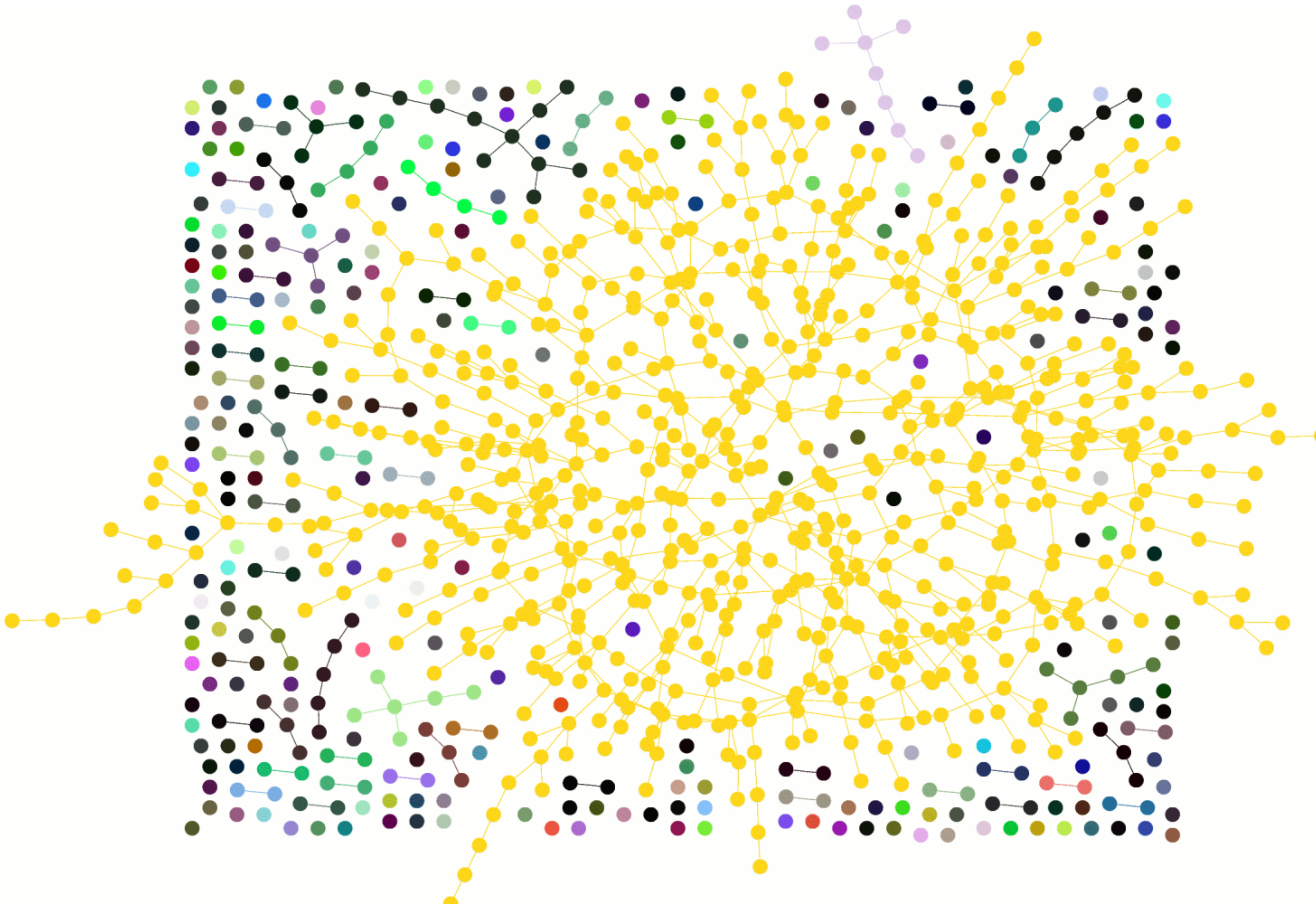
G(1000,0.2/1000)



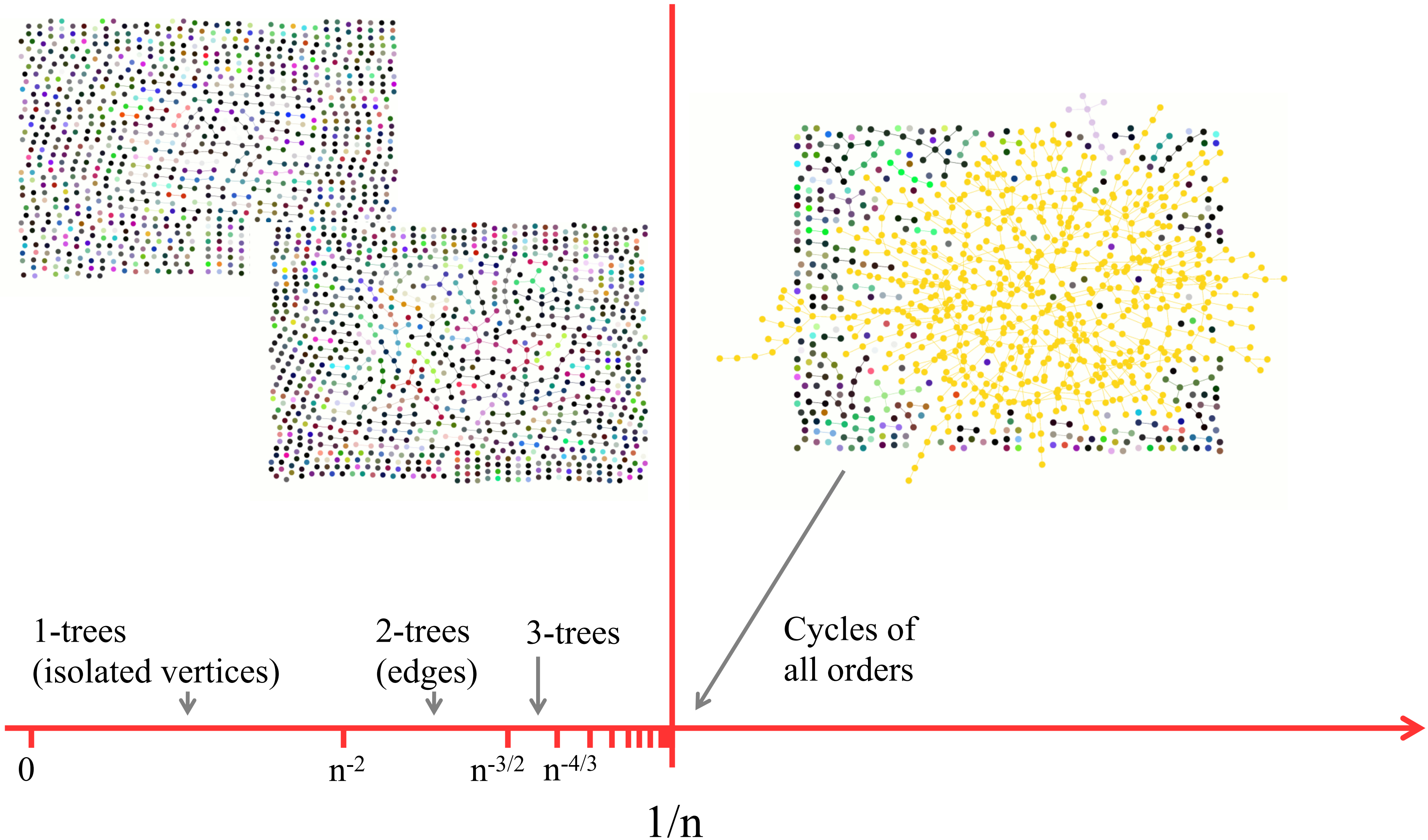
G(1000,0.5/1000)



G(1000,1.5/1000)



Recap: The Evolution from p=0 to p=1/n



The Giant Component

- Threshold function for giant component: $t(n) = 1/n$
- More precise: set $p(n) = c/n$
  - $c>1$: unique giant component
  - $c<1$: only small components
  - $c=1$: need to define even more fine scaling

Small Components for c<1

- Theorem:
  - If c<1, then the largest component of G(n,c/n) has a.a.s. at most

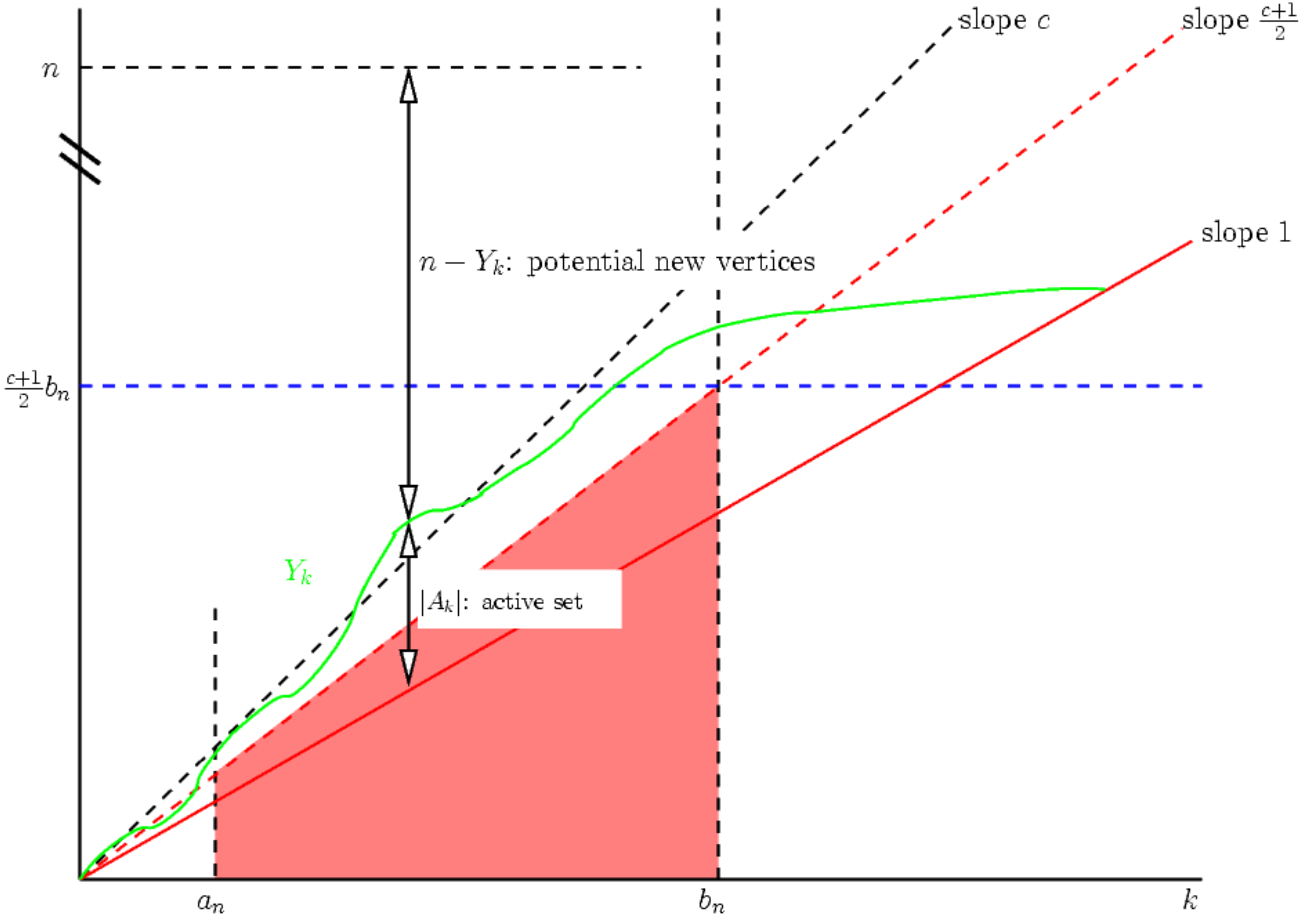
$$\frac{3}{1-c^2}\log(n)$$

  vertices
- Proof:
  - $c=1-\varepsilon$
  - Y_i: total number of vertices visited so far (saturated and active)
  - Y_i: Markov chain with $Y_{i+1}-Y_i \sim \text{Binom}(n-Y_i,p)$
  - Define Y_i^+: random walk with increments Binom(n,p)
  - $Y_i^+ \sim \text{Binom}(ni,p)$
  - Y_i^+ stochastically dominates Y_i

Large Component for c>1

- Theorem:
  - There is a unique giant component for c>1: The largest component of G(n,p) has θ(n) vertices, and the second-largest has O(log n) vertices
- Proof has three parts:
  - Part 1: each component is either small or quite large
  - Part 2: Large component is unique
  - Part 3: Large component has size θ(n)

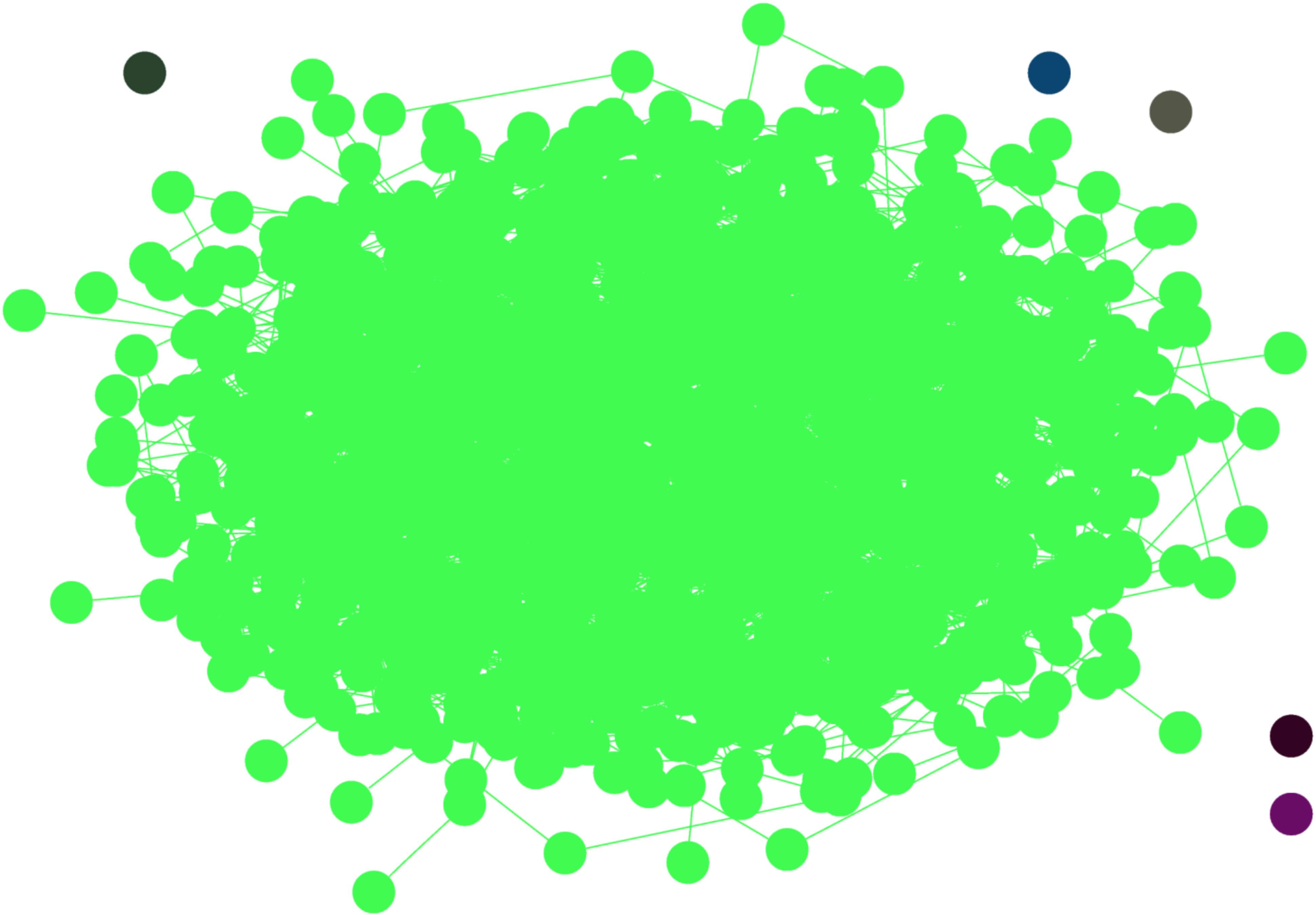
Component Exposure Process for c>1



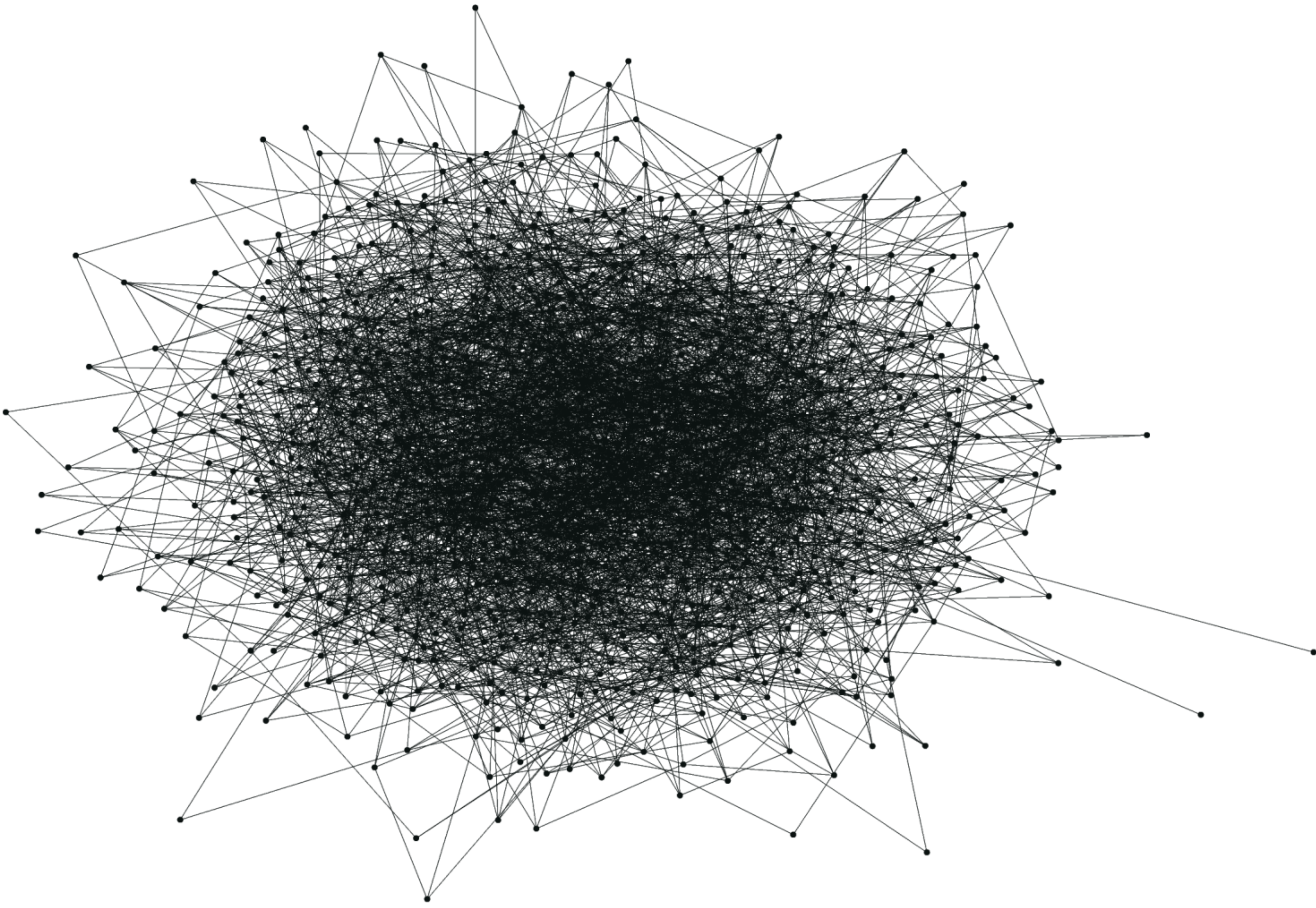
Connectivity

- Theorem:
  - $t(n) = \log n/n$ is a threshold function for the disappearance of isolated vertices
- Intuition:
  - When only a small number of isolated vertices left: P(vertex connected) = $q \ll 1$
  - P(component of size k isolated) ~ q^k

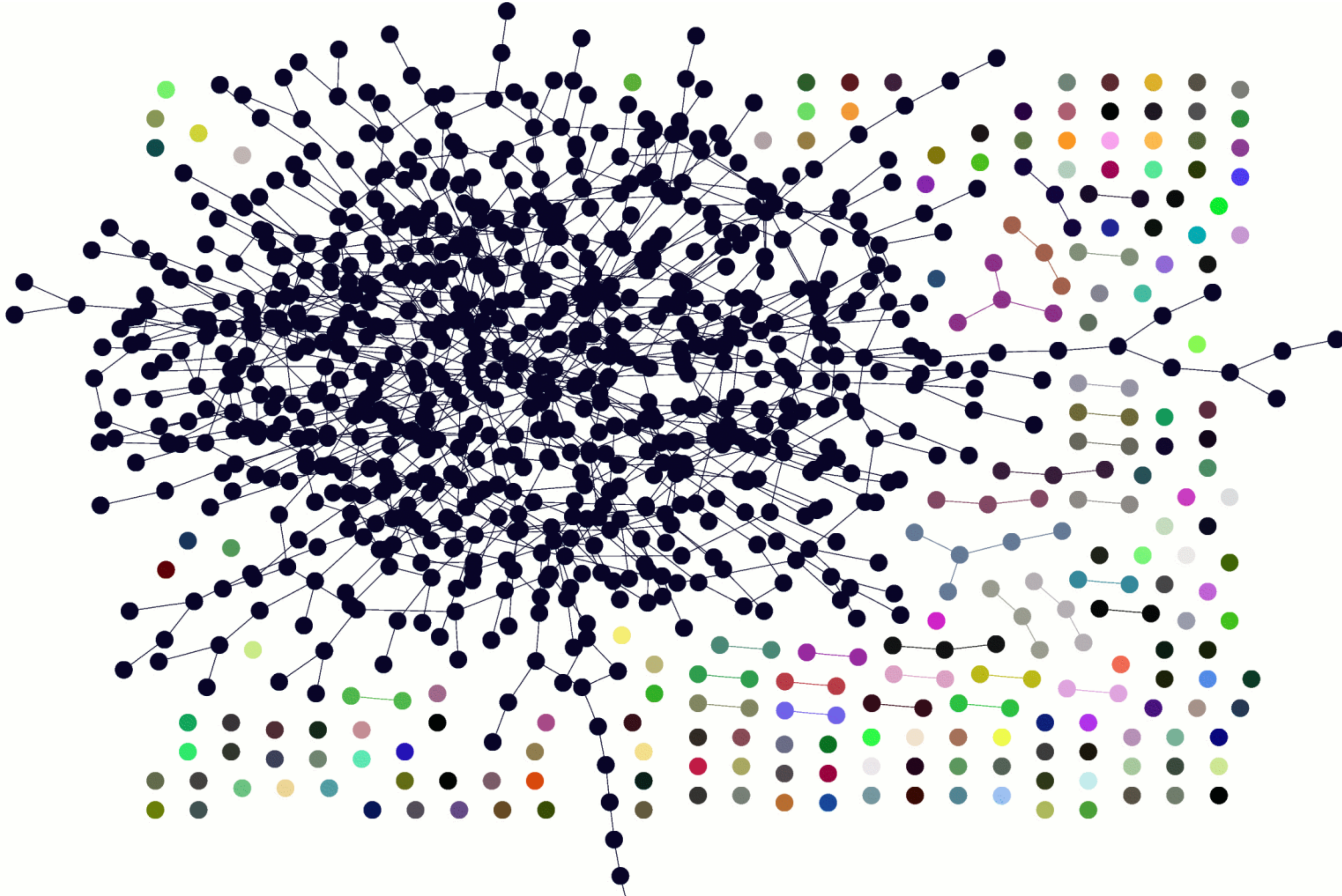
G(1000,5/1000)



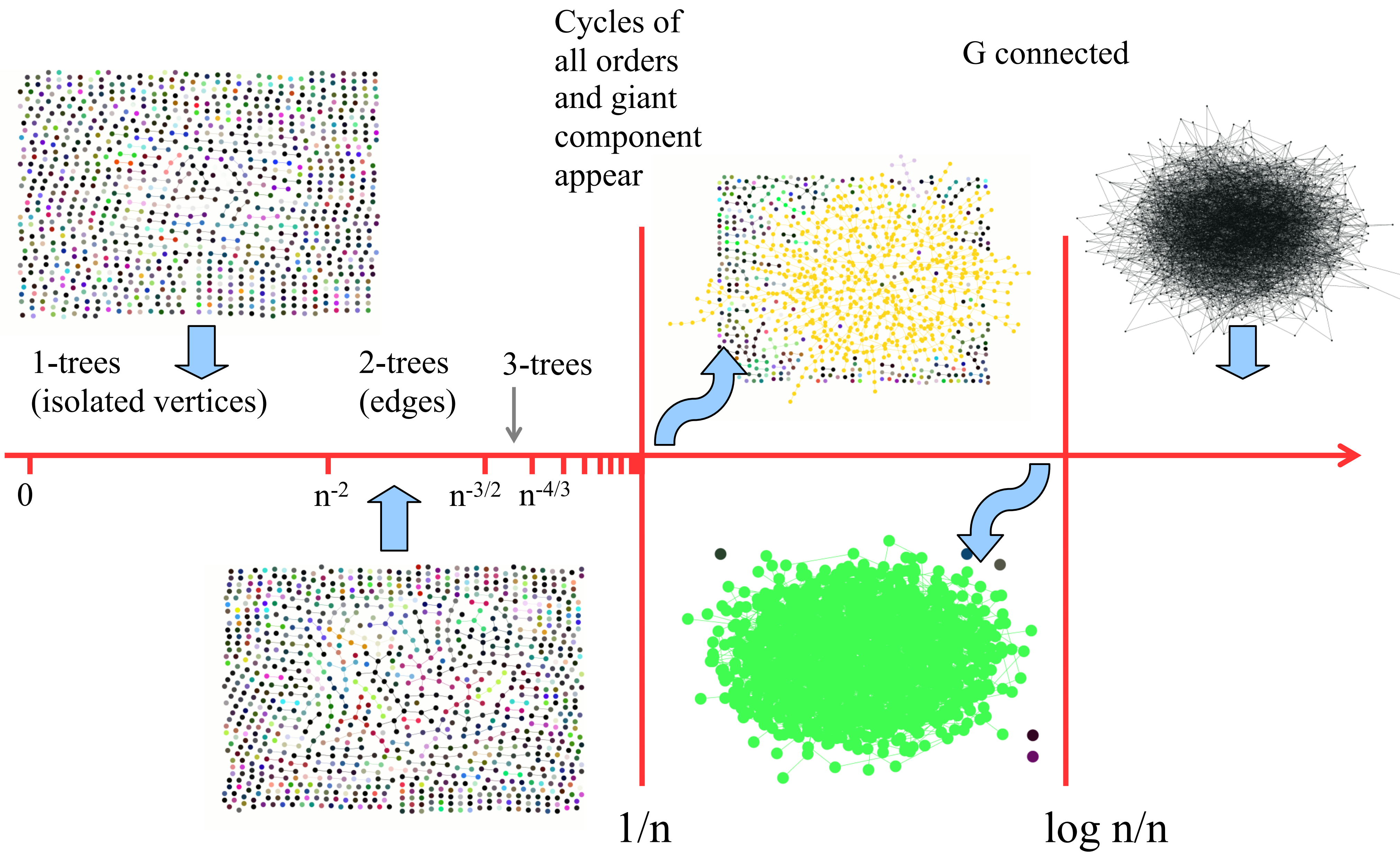
G(1000,8/1000)



G(1000,2/1000): Giant Component + Trees



Recap: The Evolution of G(n,p)

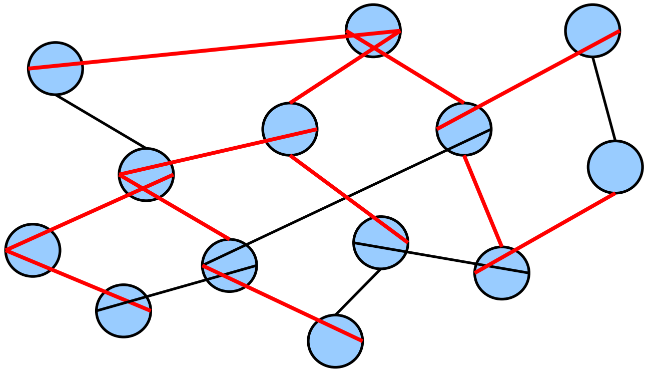


Connectivity

- Theorem:
  - $t(n) = \log n/n$ is a threshold function for the disappearance of isolated vertices
- Intuition:
  - When only a small number of isolated vertices left: P(vertex connected) = $q << 1$
  - P(component of size k isolated) ~ q^k

Connectivity (cont.)

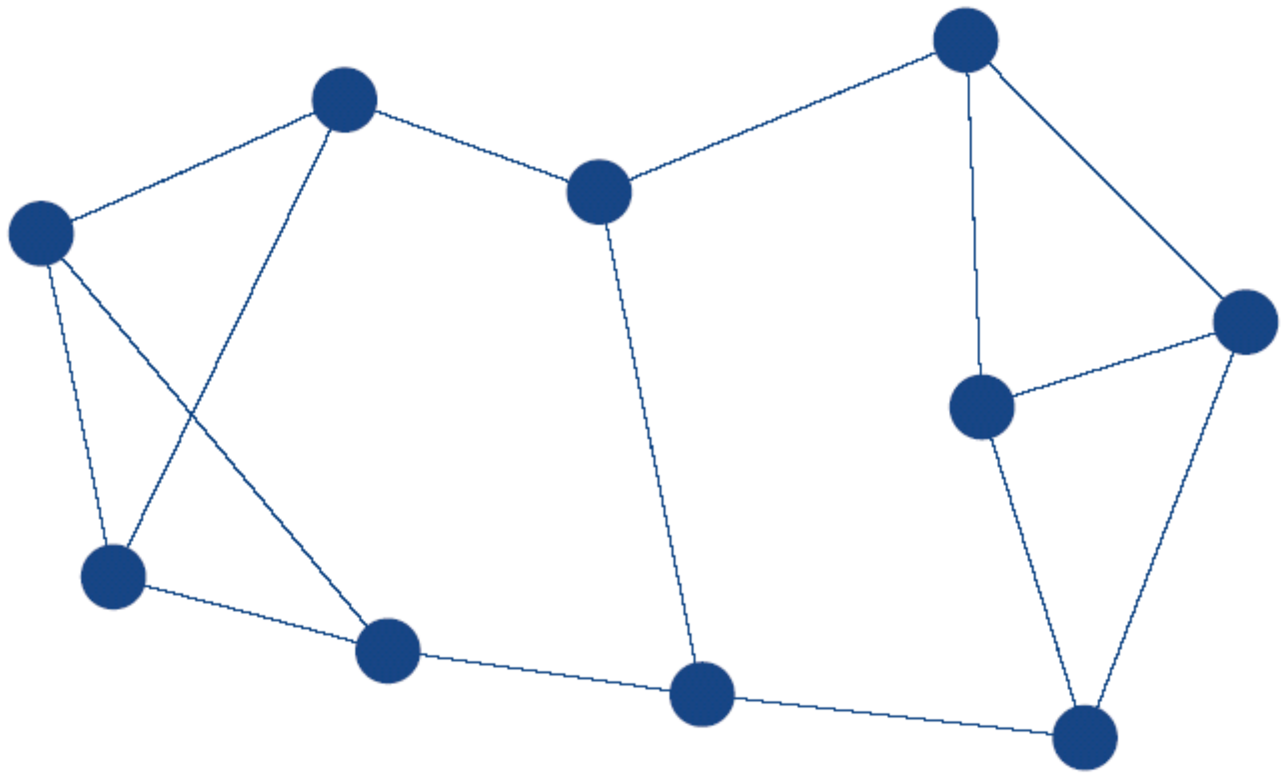
- Theorem:
  - t(n) = log n/n is a threshold function for connectivity
- Proof:
  - Show that P(component of size <n/2 appears) is small
  - Cayley's formula: # labeled trees of order k = k^{k-2}



$$
\begin{aligned}
&\mathbb{P}\{G(n,p) \text{ contains component of order } k\} \\
&\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\
&\leq k^{-2} \exp\left[k(\log n + 1) + (k-1)(\log(\omega_n \log n) - \log n) - k\omega_n \log n + \frac{k^2}{n}\omega_n \log n\right] \\
&\leq nk^{-2} \exp\left[k + k\log(\omega_n \log n) - \frac{1}{2}k\omega_n \log n\right] \\
&\leq nk^{-2} \exp\left[-(1/3)\omega_n k \log n\right] \qquad (\text{for } n \text{ large enough}) \\
&= k^{-2} n^{1-k\omega_n/3},
\end{aligned}
$$

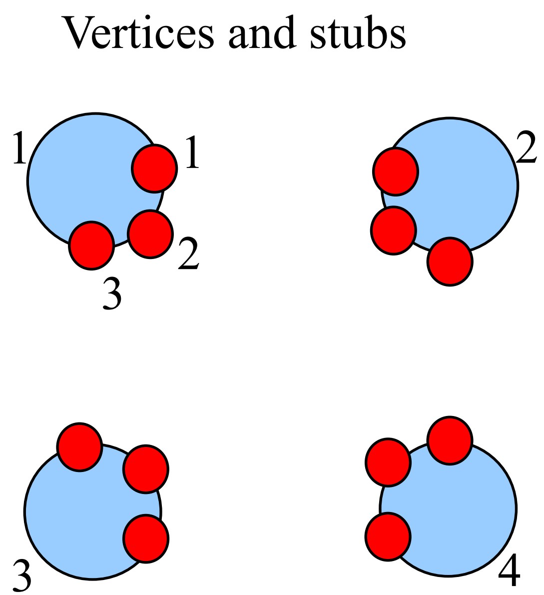
Random Regular Graph G(n,r)

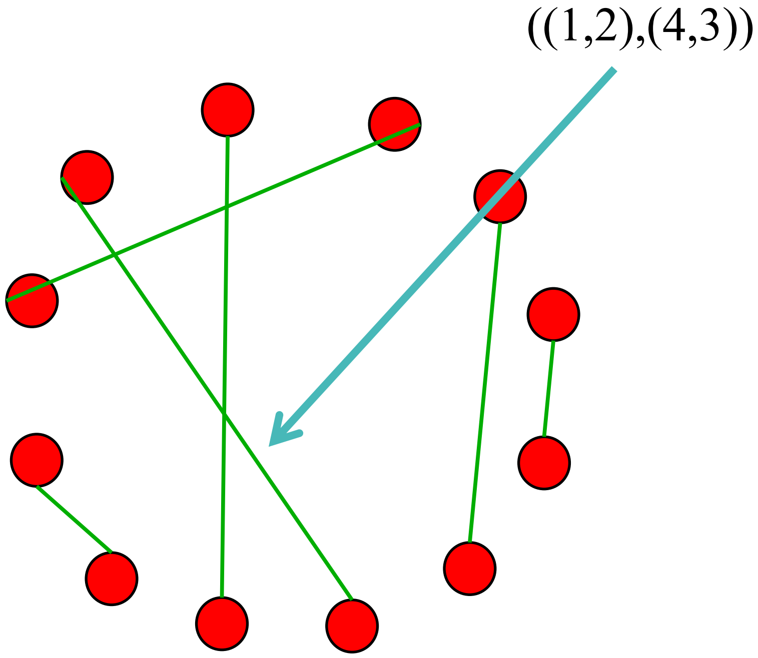
- G(n,r) is uniformly sampled from set of all graphs of order n and constant degree r
- Note: edges dependent -> probability space harder to describe
  - Detour: simpler model that generates G(n,r) with nonzero probability



The Pairing Model

- Pairing model:
  - r labeled stubs (half-edge) per vertex

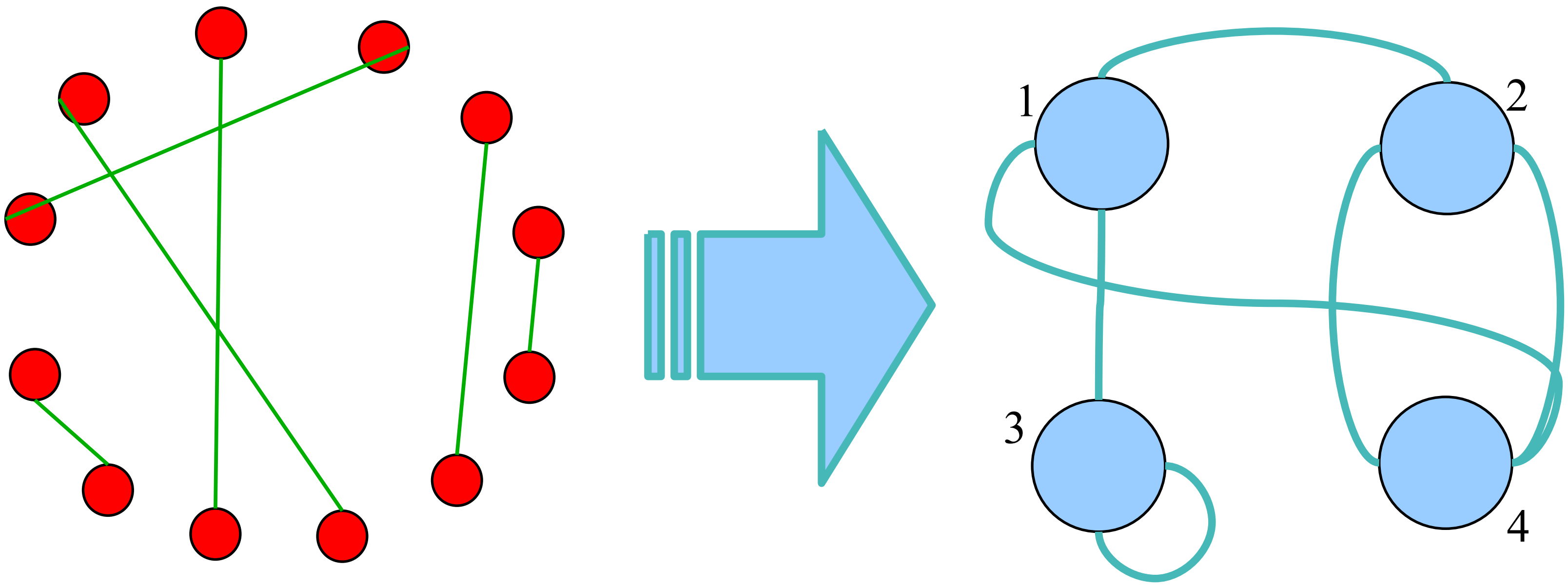




The Pairing Model (cont.)

Random matching

G*(n,r): random regular multigraph



Projection:
forget the edge labels

- Note:
  - # pairings = (nr-1)!! = (nr-1)(nr-3)...3

Appearance of H in G*(n,r)

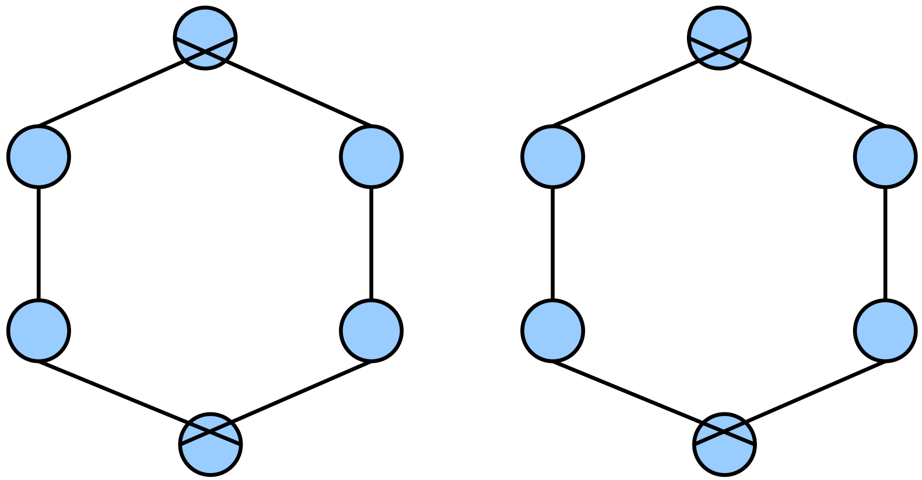
- Theorem:
  - Z_k = # of k-cycles in G*(n,r)
  - Random variables $\{Z_k\}$, k>=1 converge in distribution to independent random variables Po($(r-1)^k/2k$)
- Proof:
  - View G*(n,r) as projection of pairing model
  - Probability p_k that set of k labeled edges is in a random pairing:

$$p_k = \frac{(rn-2k-1)!!}{(rn-1)!!} = \frac{1}{(rn-1)(rn-3)\dots(rn-2k+1)}$$

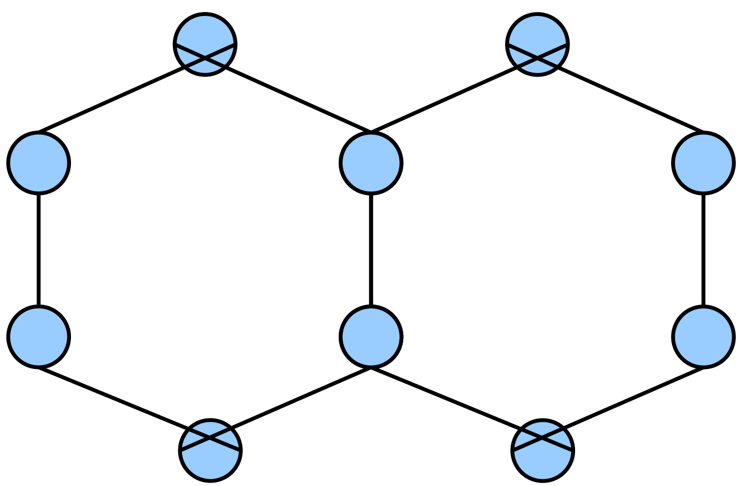
  - Show convergence of all moments

Appearance of H in G*(n,r) (cont.)

- Union of cycles



v=e



v<e

Appearance of H in G*(n,r) (cont.)

- $E[Z_k]$: expected number of k-cycles
- $E[(Z_k)_2]$: expected number of ordered pairs of distinct k-cycles
  - $E[(Z_k)_2]=S_0+S_>$
    - $S_>$=sum of terms $S_{v,e}$
    - v: # vertices in intersection
    - e: # edges in intersection
    - Number of terms $S_{v,e}$ does not depend on n
  - S_0:

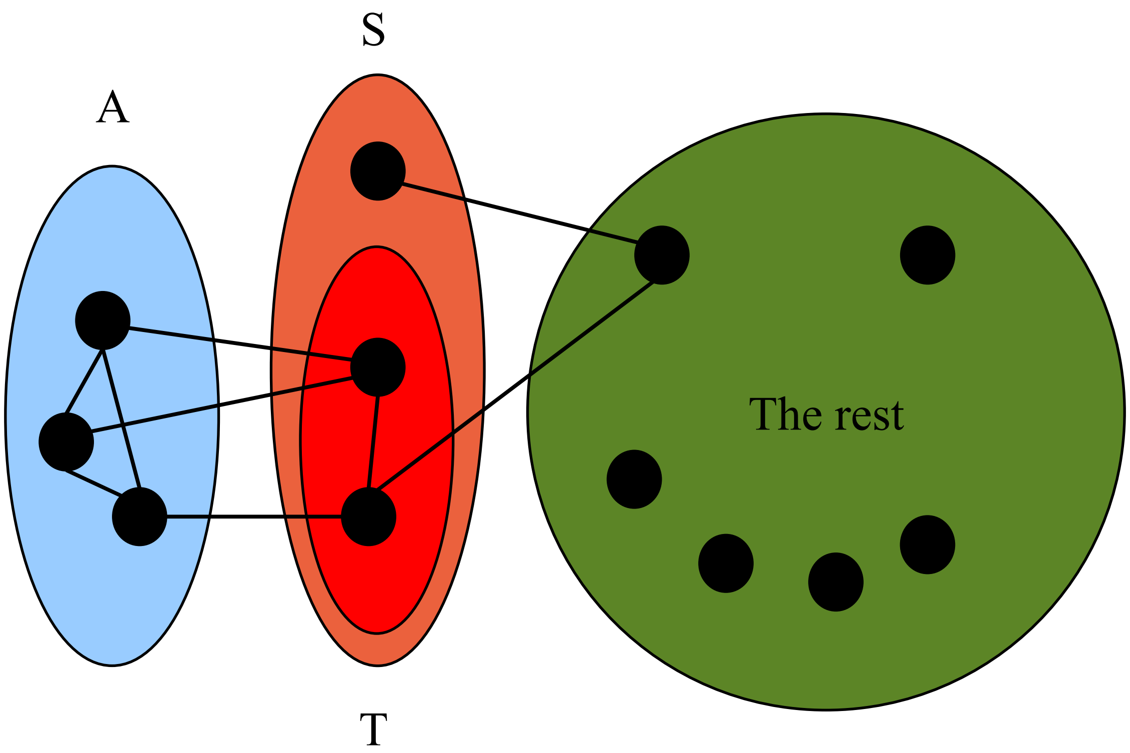
$$\binom{n}{k}\frac{k!}{2k}\binom{n-k}{k}\frac{k!}{2k}(r(r-1))^{2k}=\left(\frac{n^k r^k (r-1)^k}{2k}\right)^2$$

The Random Regular Graph G(n,r) (finally!)

- Theorem:
  - The random variables (Z_k) converge in distribution to a collection of independent $Po((r-1)^k/2k)$
- Theorem:
  - $P(G(n,r)$ is simple$) = \exp(-(r^2-1)/4)$
- Proof:
  - $P(G$ is simple$) = P(Z_1=Z_2=0)$
- Theorem:
  - Any a.a.s. property for G*(n,r) is also an a.a.s. Property of G(n,r); the converse is false

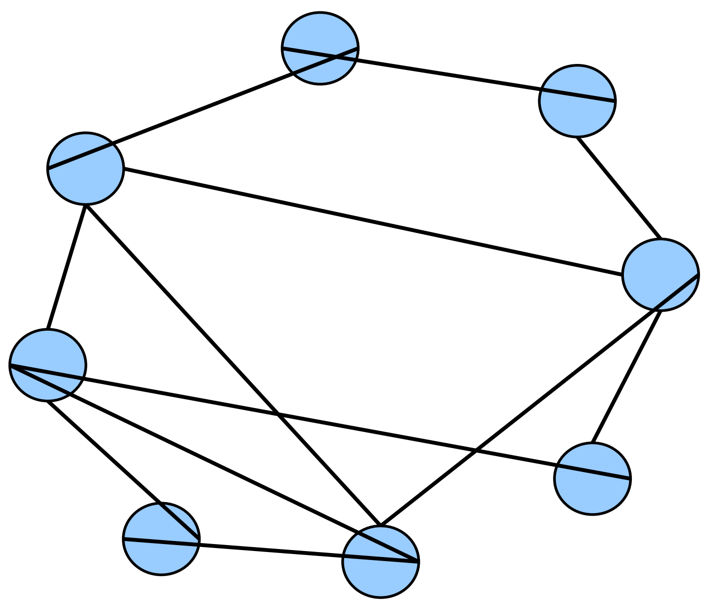
Connectivity of G(n,r)

- Theorem:
  - For r > 2, G(n,r) is connected a.a.s.
- Proof:



G(n,D): Generalized Degree Distribution

- Model for generalized degree distribution
  - $d_i(n)$: number of vertices of degree i
  - $d_i(n)/n \to \lambda_i$
- Generate G(n,D) through generalized version of pairing model



Condition for Giant Cluster in G(n,D)

- Theorem:

$$Q(D) = \sum i\,(i-2)\lambda_i$$

  - If Q(D) > 0, then there is a unique giant cluster
  - If Q(D) < 0, then the largest cluster is O(log n)