Chasing the k-colorability threshold

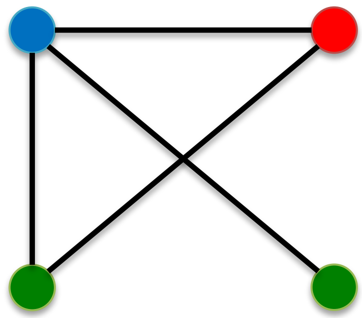
Danny Vilenchik

The Weizmann Institute

Join work with Amin Coja-Oghlan

Appeared in FOCS 2013

Graph colorability problem



$\chi(G) = 3$

- Worse case: NP-hard and no approximation

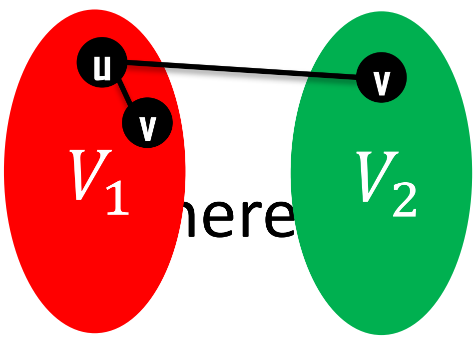
What will typically be the chromatic number of the random graph $G(n, m)$?

What should expect?

Z_k counting proper k-colorings of $G(n, m = dn/2)$

$$E[Z_k] = \sum_{\sigma} \Pr[\sigma\ is\ k - coloring]$$

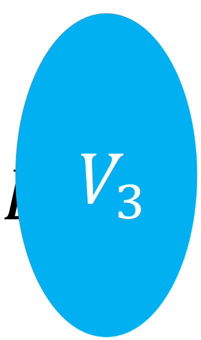
$$\Pr[\sigma\ is\ a\ k - coloring] \approx \left(1 - \frac{1}{k}\right)^m$$



here

Today: Assume σ is balanced

$E[Z_k]$



$\to 0$ at $d_{first} \approx 2k \ln k$ ⇒ $\chi(G) \approx d/(2 \ln d)$

Pre-History

- Bollobas (88) Luczak (91) - right asymptotic for $m \gg n$
- Shamir & Spencer proved $O(1)$ −concentration $m \ll n^{3/2}$
- Two-point concentration by Alon and Krivelevich in 1997

One point concentration?

Achlioptas & Naor

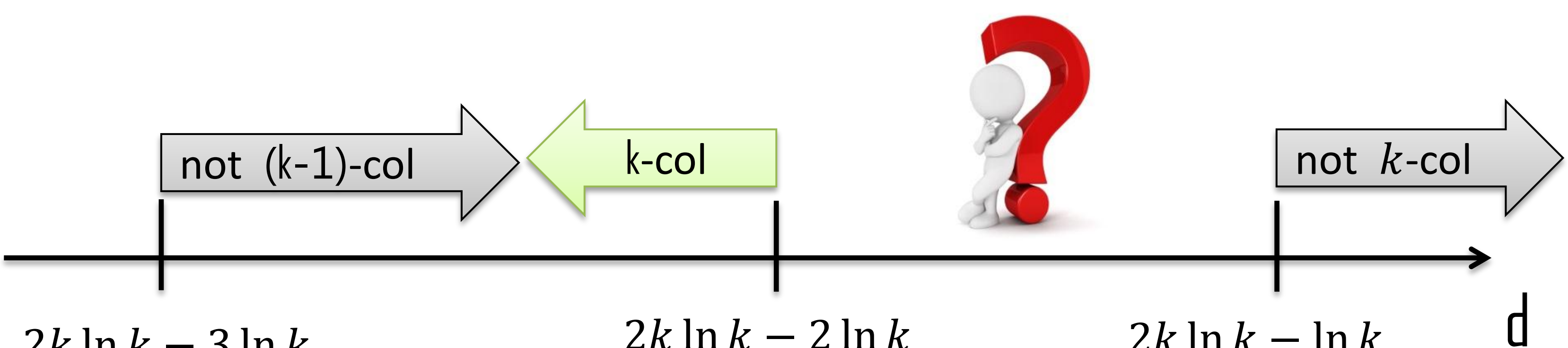
[Annals of Mathematics 2005]

For fixed large k, the k-colorability threshold satisfies

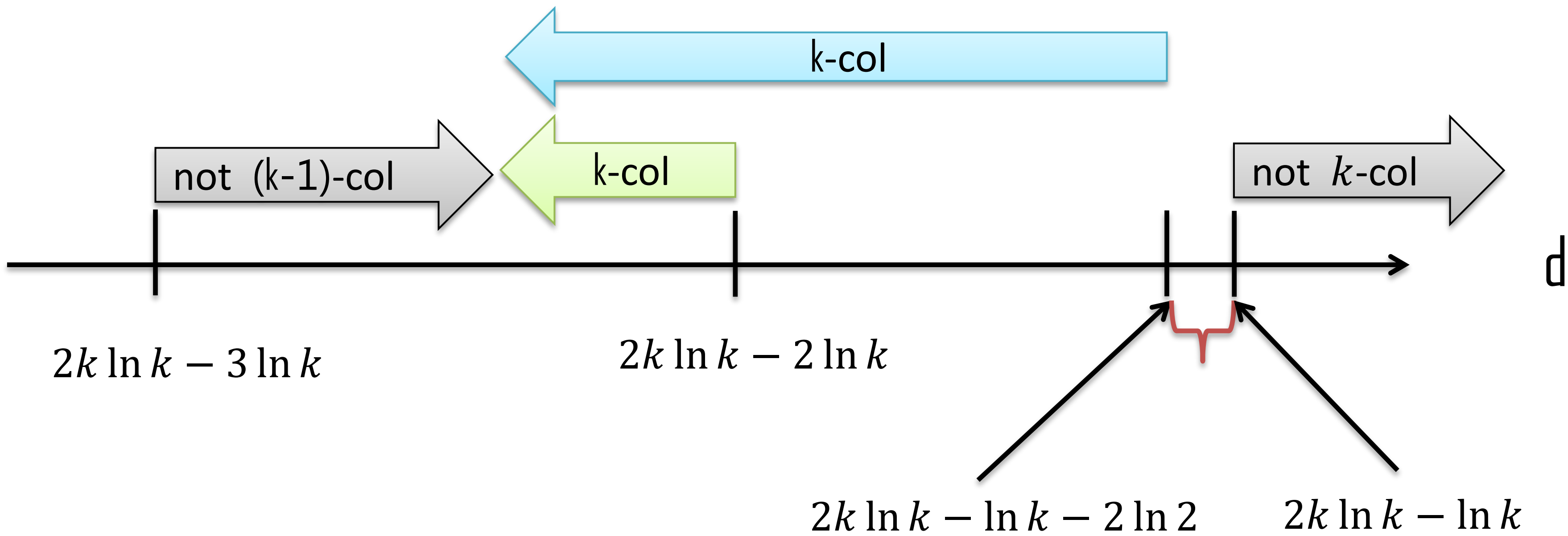
$$d_{k-col} \geq 2k \ln k - 2 \ln k - 2$$

$$+$$

$$d_{k-col} \leq d_{first} = 2k \ln k - \ln k$$



Our Result



We determine the exact chromatic number for all but a vanishing (with k) fraction of densities d

Physicist predict $2k \ln k - \ln k - 1$ (Cavity Method)

The Second Moment Method

- Achlioptas & Peres 2003 (k-SAT)
- Coja-Oghlan & Panagiotou 2012-13 (k-SAT, NAE k-SAT)
- Coja-Oghlan & Zdeborova 2012: Hypergraph 2-col.
- Dyer, Frieze and Greenhill 2014: Chromatic number of random hypergraphs
- Coja-Oghlan, Efthymiou, Hetterich 2014: Chromatic number of d-regular graphs

The Second Moment Method

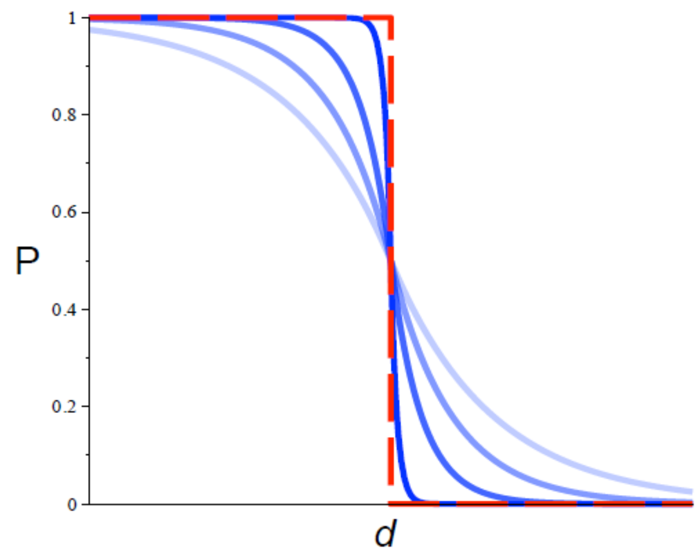
- $E[Z_k] > 0$ doesn't imply $Z_k > 0$ whp
- Need to control the variance (second moment)
- Suppose that $Z(G)$ is such that

$$Z > 0 \rightarrow G \text{ is k-colorable}$$

- Further suppose that $0 < E[Z^2] \leq C \cdot E[Z]^2$
- The Paley-Zygmund inequality says that

$$\Pr[Z > 0] > \frac{E[Z]^2}{E[Z^2]} \geq \frac{1}{C}$$

Achlioptas and Friedgut 99: For any fixed $k \geq 3$ there is a sharp threshold sequence $d_{k-col}(n)$



Use Sharp Threhold + PZ to get a lower bound on d_{k-col}

We "just" need to find a random variable Z so that

- $Z > 0 \rightarrow G$ is k-colorable
- $0 < E[Z^2] \leq C \cdot E[Z]^2$

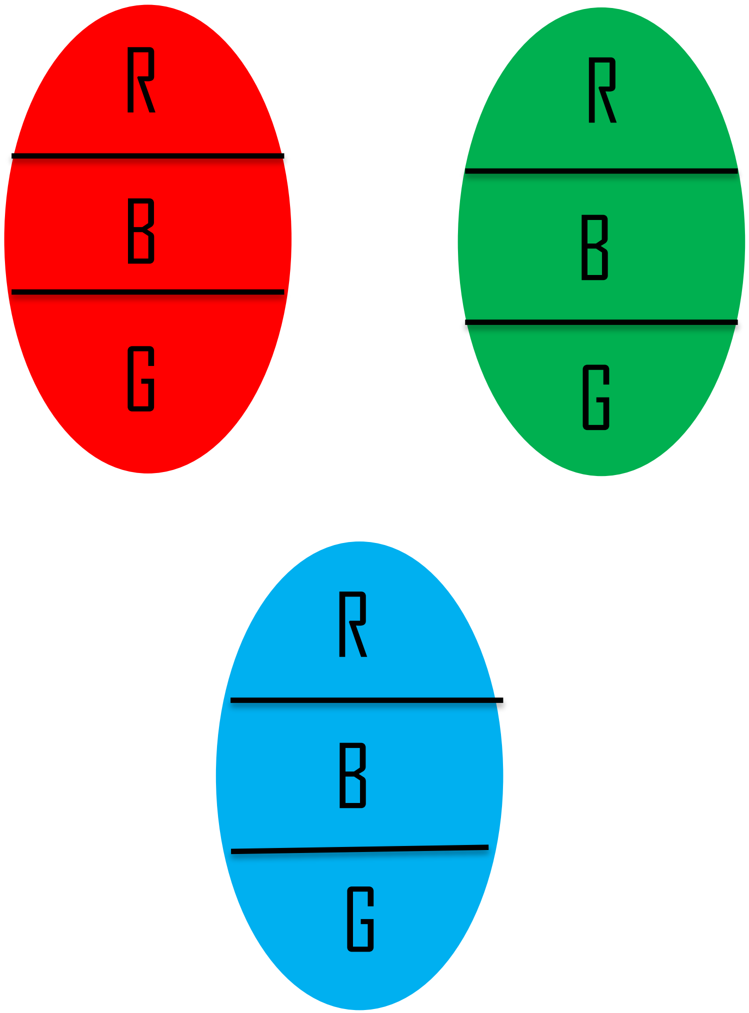
Counting Pairs of Colorings

- Goal: Compute $E[Z^2]$ - the expected number of pairs of colorings

$$\Pr[\sigma \text{ and } \tau \text{ are both proper k-col of } G]$$

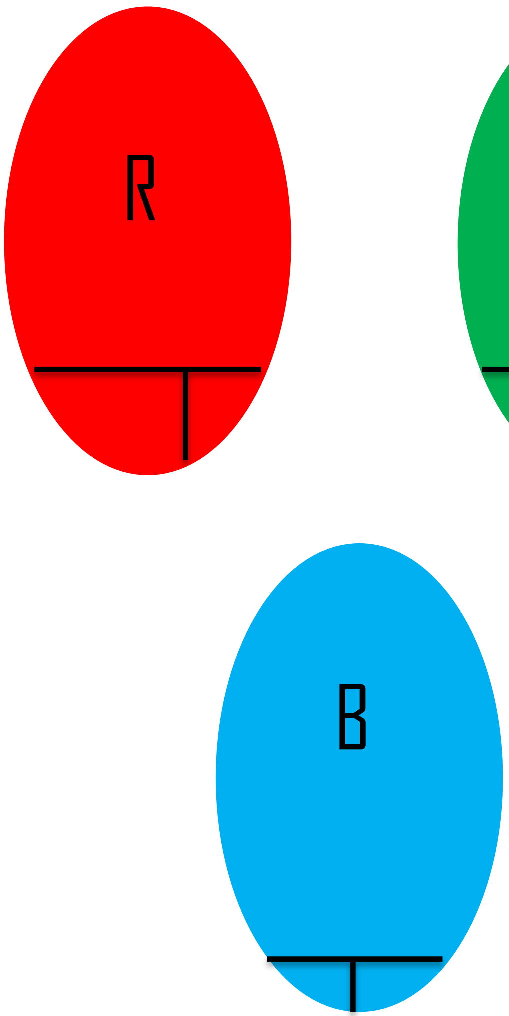
- This quantity depends on the "distance" between σ and τ
  - Overlap matrix ρ is a $k \times k$ matrix where

Overlap Matrix Example



$$\rho^* = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Overlap Matrix Example



$$\begin{array}{cc} & \begin{array}{ccc} R & G & B \end{array} \\ \rho = & \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.15 & 0.8 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix} \end{array}$$

First try – Balanced Colorings

- $Z_{k,bal}$ - the number of balanced k-colorings
- $Z_{\rho,bal}$ - number of coloring pairs (σ, τ) that obey ρ

$$E\left[{Z_{k,bal}}^2\right] = \sum_{balanced\ \rho} E[Z_{\rho,bal}]$$

Probability of Pairs

A_σ = a random edge is monochromatic under σ

$$\Pr[\sigma \text{ and } \tau \text{ are proper k-col}] = (1 - \Pr[A_\sigma \vee A_\tau])^m =$$

$$= (1 - \Pr[A_\sigma] - \Pr[A_\tau] + \Pr[A_\sigma \wedge A_\tau])^m =$$

$$= \left(1 - \frac{2}{k} + \frac{\|\rho\|^2}{k^2}\right)^m$$

$$\|\rho\|^2 = \sum_{1 \le i,j \le k} (\rho_{i,j})^2 \quad \text{(Frobenius norm)}$$

Balanced k-colorings contd.

$$E\left[{Z_{k,bal}}^2\right] = \sum_{\rho} \underbrace{k^n \binom{n}{\rho_{11}\frac{n}{k}, \rho_{12}\frac{n}{k}, \ldots, \rho_{kk}\frac{n}{k}} \left(1 - \frac{2}{k} + \frac{\|\rho\|^2}{k^2}\right)^m}_{f(\rho)}$$

Our goal is to prove that this is $\leq C(k)(E[Z_{k,bal}])^2$

Let ρ^* be the matrix with all entries equal $1/k$

$$f(\rho^*) \approx E[Z_{k,bal}]^2$$

Suffices to show that ρ^* is the maximizer

Back to Achlioptas and Naor

- AN relax by maximizing over singly-stochastic matrices
  - This reduces the dimension from k^2 to k
- Their analysis is complicated and non-combinatorial
- They manage to prove that $f(\rho^*)$ is the max up to

$$d_{AN} = 2k\ln k - 2\ln k - 2$$

What goes wrong beyond d_{AN}?

- Max is at a matrix which is not doubly-stochastic
- So is it just the relaxation?

No …

$$\rho' = \left(1 - \frac{1}{k}\right) I_k + \frac{1}{k^2} J_k$$

For some $d_{AN} < d < d_{first}$

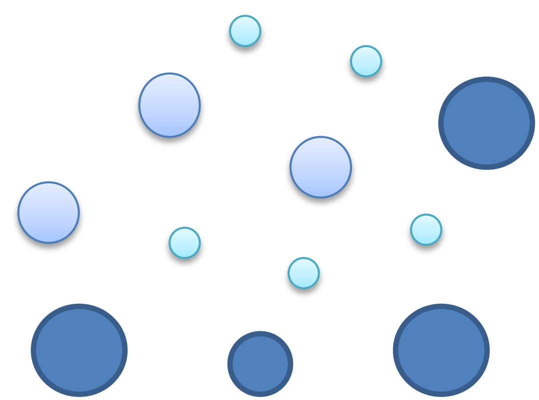
$$f(\rho') > f(\rho^*)$$

Our approach: a new R.V.

- Cluster of σ

$$C(\sigma) = \{\tau : \forall i \ \rho_{ii}(\tau, \sigma) \geq 0.99\}$$

- We say that σ is good if
  - There is no τ s.t. $\rho_{ii}(\tau, \sigma) \in [0.51, 0.99]$
  - $|C(\sigma)| \leq E[Z_k]$
- We then consider the r.v. $Z_{k,good}$



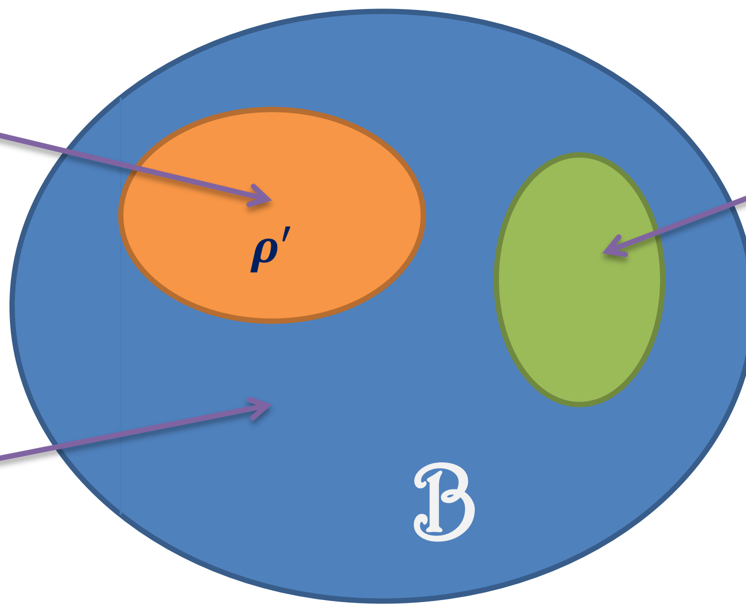
The New 2nd Moment

Goal: $E\left[{Z_{k,good}}^2\right] \leq C(k)\left(E\left[Z_{k,good}\right]\right)^2$

$$E\left[{Z_{k,good}}^2\right] = \sum_{\rho} E\left[Z_{\rho,good}\right]$$

Some color has overlap ≤ 0.51

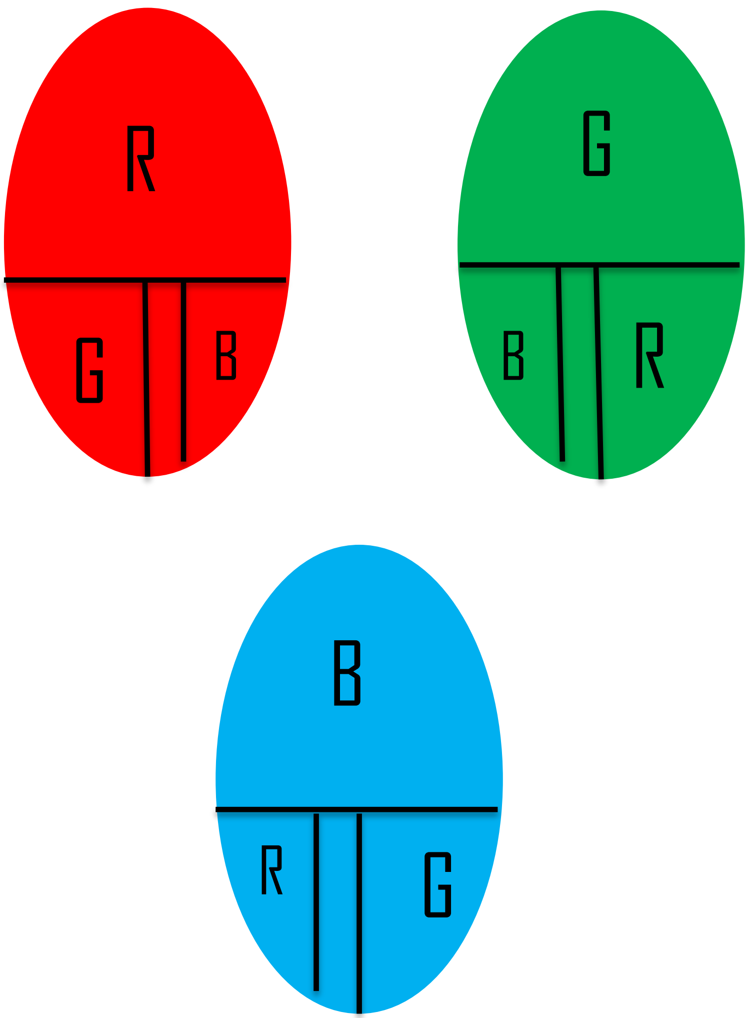
All color classes are have large (>0.999) overlap



Bad overlap matrices - Ignored

$$\rho' = \left(1 - \frac{1}{k}\right) I_k + \frac{1}{k^2} J_k$$

Pairs From Different Clusters



$$\rho = \begin{array}{c} \begin{array}{ccc} R & G & B \end{array} \\ \begin{pmatrix} 0.4 & 0.35 & 0.25 \\ 0.35 & 0.4 & 0.25 \\ 0.25 & 0.35 & 0.4 \end{pmatrix} \end{array}$$

$$\rho' = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

$$f(\rho') > f(\rho)$$

2^{nd} moment contd.

$$E\left[{Z_{k,good}}^2\right] = \sum_{\rho} E[Z_{\rho,good}]$$

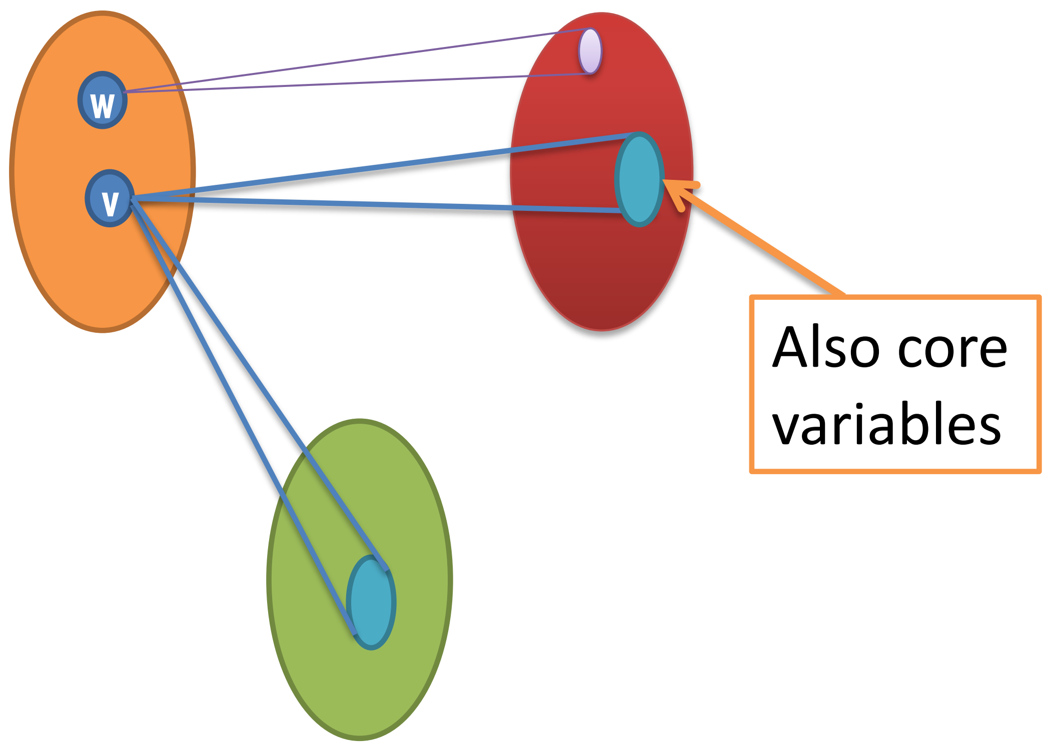
Key Observation: If two good colorings obey ρ then they have the same clusters

$$\sum_{\rho} E\left[Z_{\rho,good}\right] = k! \sum_{\sigma} E[C(\sigma)|\sigma\ is\ good] \Pr[\sigma\ is\ good] =$$

$$k! \max_{\sigma} E[C(\sigma)|\sigma\ is\ good] \sum_{\sigma} \Pr[\sigma\ is\ good] \leq k!\,(E[Z_k])^2$$

Are there any good colorings?

- We show that w.h.p. a random k-coloring σ is good
  - We define the notion of core and free variables



v belongs to the core

w is 1-free

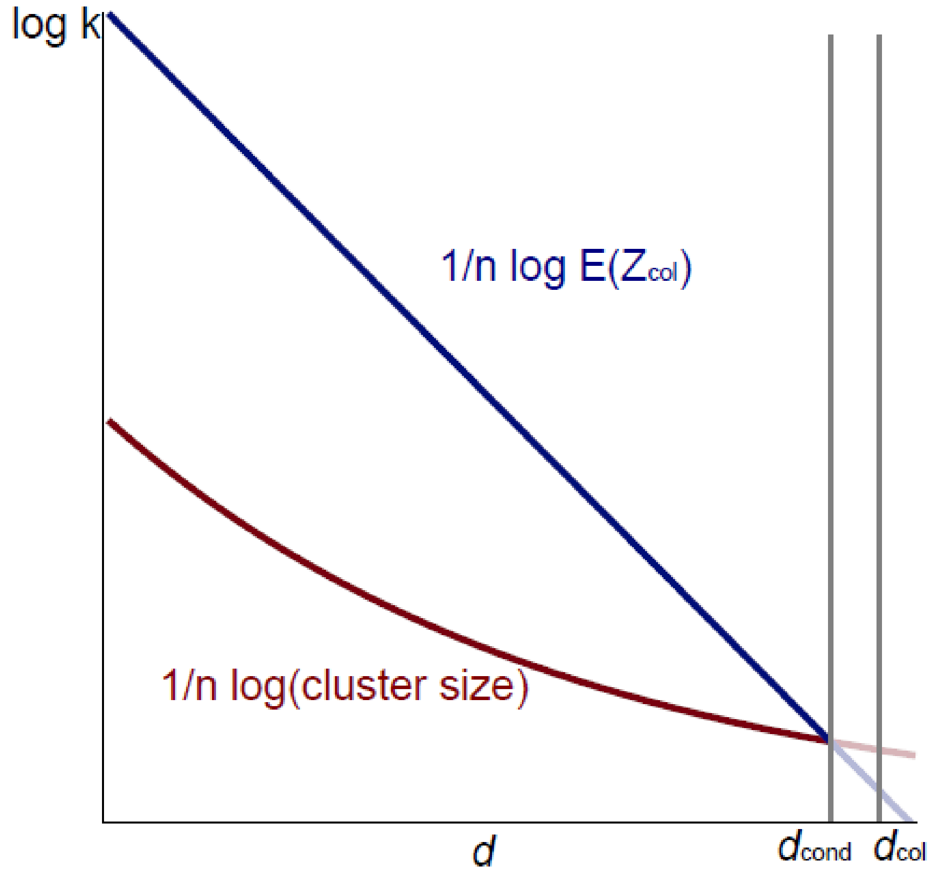
Are there good colorings?

- Use expansion to show that every τ either
  - Agrees with σ on the core
  - Or, is far away from σ
- We show that $|C(\sigma)|$ satisfies

$$|C(\sigma)| \leq 2^{\#1-free\ variables} \leq E[Z_k]$$

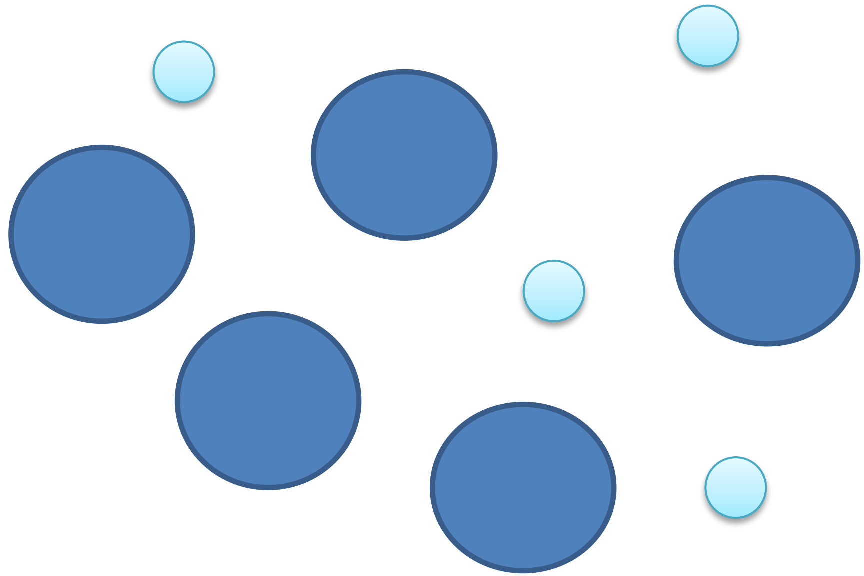
This property is violated above the condensation threshold

Condensation



The functions cannot extend analytically beyond d_{cond}

Condensation Cont.



- Few large clusters dominate the topology
- A typical pair of colorings will NOT be uncorrelated
- ρ^* (matrix with all entries equal $1/k$) will not be the max!