Solution geometry of a random k-XORSAT near the clustering threshold

Jane Gao

University of Waterloo

Monash University – Discrete Maths Seminar

June 30, 2014

Collaborator: Mike Molloy

A few words about random structures

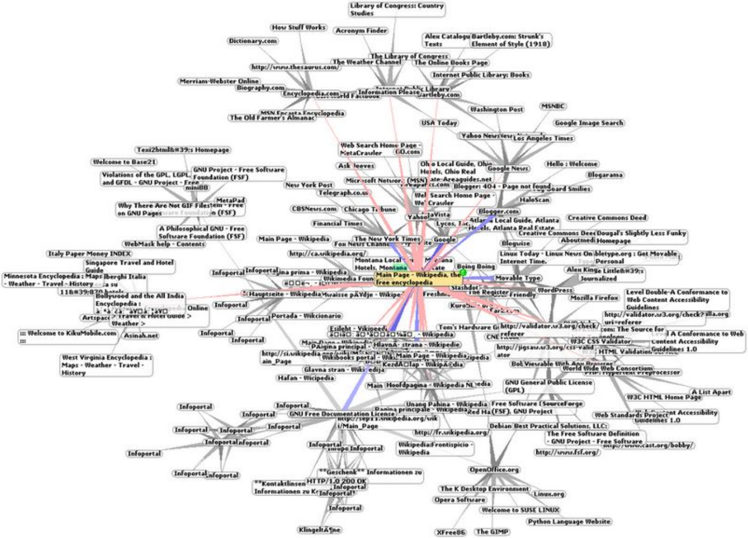
- What are random structures?
- Why study random strcutures?

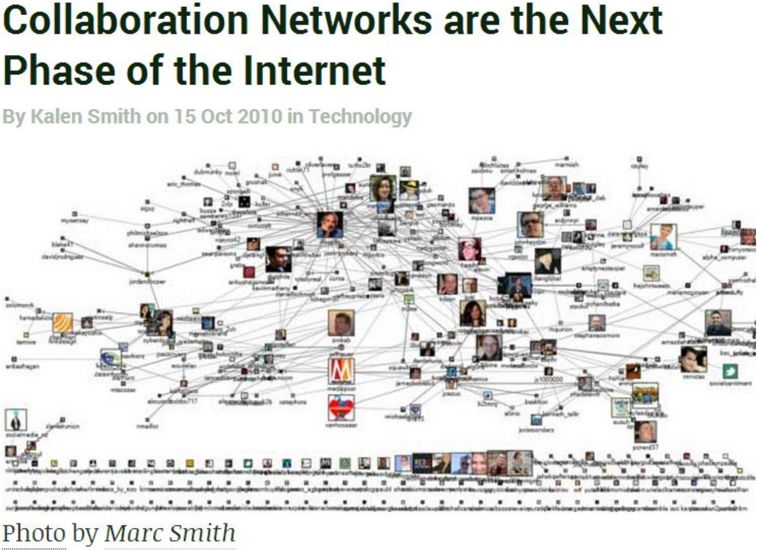
A few words about random structures

- What are random structures?
- Why study random strcutures?

A few words about random structures

- What are random structures?
- Why study random strcutures?

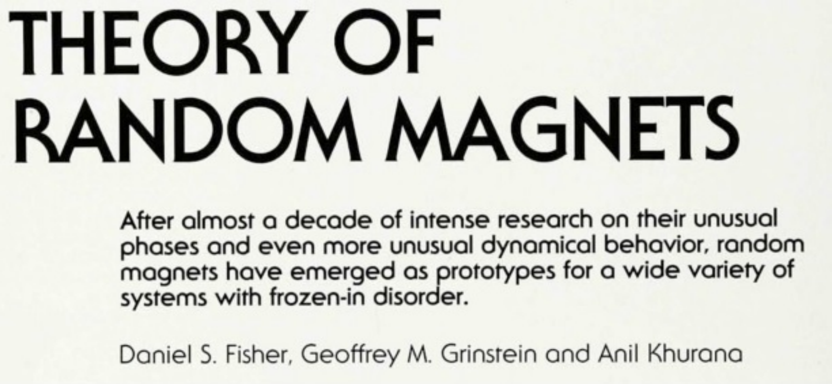

Library of Congress: Country Studies
How Stuff Works
Acronym Finder
Dictionary.com
Alex Catalogue
Texts
Bartleby.com: Strunk's Element of Style (1918)
The Library of Congress
The Weather Channel
The Online Books Page
http://www.thesaurus.com/
Internet Public Library: Books
Merriam-Webster Online
Biography.com
Encyclopedia.com
Information Please
Bartleby.com
Washington Post
MSNBC
MSN Encarta Encyclopedia
The Old Farmer's Almanac
USA Today
Google Image Search
Yahoo News
Los Angeles Times
Text2html's Homepage
Welcome to Base21
Web Search Home Page - MetaCrawler
GO.com
Ask Jeeves
Ohio Local Guide, Ohio Hotels, Ohio Real
Google News
Hello : Welcome
Blogarama
GNU Project - Free Software Foundation (FSF)
Violations of the GPL, LGPL, and GFDL - GNU Project - Free
miniBB
New York Post
Microsoft Network (MSN)
Blogger: 404 - Page not found
Telegraph.co.uk
Web Search Home Page - Web Crawler
HaloScan
MetaPad
Why There Are Not GIF Files on GNU Pages
CBSNews.com
Chicago Tribune
Blogger.com
Yahoo!
AltaVista
Atlanta Local Guide, Atlanta Hotels, Atlanta Real Estate
Creative Commons Deed
A Philosophical GNU - Free Software Foundation (FSF)
Financial Times
Lycos, Inc.
Creative Commons Deed
Dougal's Slightly Less Funky Homepage
WebMask help - Contents
Main Page - Wikipedia
The New York Times
Fox News Channel
Google
Aboutmedia
Blogwise
Italy Paper Money INDEX
http://ca.wikipedia.org/
Linux Today - Linux News On Internet Time.
Movable Type
Singapore Travel and Hotel Guide
Montana Local Guide, Hotels, Montana
Boing Boing
Personal
Minnesota Encyclopedia : Maps - Weather - Travel - History
Alberghi Italia
Infoportal
Main Page - Wikipedia, the free encyclopedia
Wikimedia Foundation
Alex King
Little's Journalized
Slashdot
The Register
Friendly
WordPress
Bollywood and the All India Encyclopedia
Hauptseite - Wikipedia
Freshmeat
Mozilla Firefox
Level Double-A Conformance to Web Content Accessibility Guidelines
Online
Infoportal
Artspace
Travel & Hotel Guide
Weather
Portada - Wikcionario
Kuro5hin
http://validator.w3.org/check?uri=referer
Infoportal
Welcome to KikuMobile.com
Asinah.net
Infoportal
Tom's Hardware Guide
http://validator.w3.org/check/referer
CNET
A Conformance to Web Content Accessibility Guidelines 1.0
W3C CSS Validator
HTML Validation Service
Infoportal
Hlavná strana - Wikipédia
Main Page - Wikipedia
http://jigsaw.w3.org/css-validator
West Virginia Encyclopedia : Maps - Weather - Travel - History
http://si.wikipedia.org/wiki/Main_Page
Wikibooks portal
Bob Viewable With Any Browser
World Wide Web Consortium
PHP: Hypertext Preprocessor
Glavna stran - Wikipedija
Hafan - Wicipedia
Hoofdpagina - Wikipedia NL
GNU General Public License (GPL)
A List Apart
W3C HTML Home Page
Infoportal
Infoportal
GNU Free Documentation License
Pagina principale - Wikipedia
Free Software (SourceForge)
Red Hat
GNU Project
Web Standards Project
Accessibility Guidelines 1.0
Infoportal
Infoportal
Infoportal
http://fr.wikipedia.org/
Debian Best Practical Solutions, LLC
The Free Software Definition - GNU Project - Free Software
Infoportal
Infoportal
Wikipedia:Frontispicio - Wikipedia
http://www.cast.org/bobby/
http://www.fsf.org/
Infoportal
Infoportal
Infoportal
Geschenk Informationen zu
OpenOffice.org
**Kontaktlinsen Informationen zu
HTTP/1.0 200 OK
Infoportal
Infoportal
The K Desktop Environment
Linux.org
Opera Software
Welcome to SUSE LINUX
Infoportal
Python Language Website
XFree86
The GIMP



Collaboration Networks are the Next Phase of the Internet

By Kalen Smith on 15 Oct 2010 in Technology

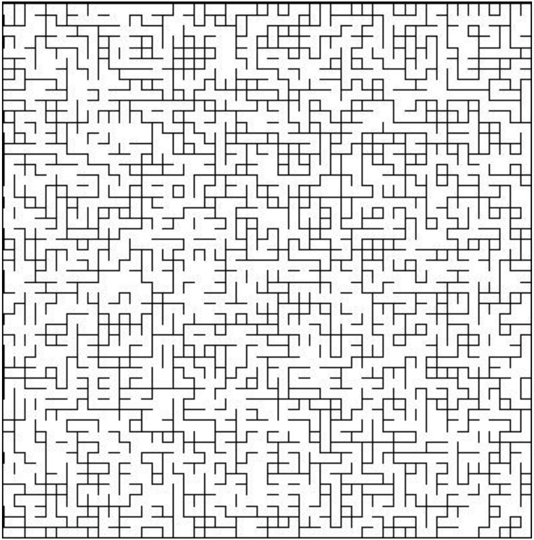
Photo by *Marc Smith*



THEORY OF RANDOM MAGNETS

After almost a decade of intense research on their unusual phases and even more unusual dynamical behavior, random magnets have emerged as prototypes for a wide variety of systems with frozen-in disorder.

Daniel S. Fisher, Geoffrey M. Grinstein and Anil Khurana



A few words about random structures

- What are random structures?
- Why study random strcutures?
- What to study in random structures?
- Evolution of random structures – phase transitions.

A few words about random structures

- What are random structures?
- Why study random strcutures?
- What to study in random structures?
- Evolution of random structures – phase transitions.

A few words about random structures

- What are random structures?
- Why study random strcutures?
- What to study in random structures?
- Evolution of random structures – phase transitions.

Phase transition; threshold

- Threshold; sharp threshold.
- "$G(n,p)$ is connected" has a sharp threshold.
- "$G(n,p)$ contains a triangle" has a (coarse) threshold.

Phase transition; threshold

- Threshold; sharp threshold.
- “$G(n, p)$ is connected” has a sharp threshold.
- “$G(n, p)$ contains a triangle” has a (coarse) threshold.

Phase transition; threshold

- Threshold; sharp threshold.
- "$G(n, p)$ is connected" has a sharp threshold.
- "$G(n, p)$ contains a triangle" has a (coarse) threshold.

Phase transition; threshold

- Threshold; sharp threshold.
- "$G(n,p)$ is connected" has a sharp threshold.
- "$G(n,p)$ contains a triangle" has a (coarse) threshold.

Random CSPs

Every clause in a k-SAT formula F (with m clauses on n literals) has the form

$$x_{i_1} \vee x_{i_2} \vee \bar{x}_{i_3} \vee \cdots \vee x_{i_k}.$$

F is satisfiable if there is a solution $(x_1, \ldots, x_n)$ such that every clause of F is satisfied.

A random instance F with m clauses and n literals is chosen uniformly at random from the set of $(2n)^{km}$ formulas.

More CSPs:

k-NAESAT, k-XORSAT, k-COL of graphs, 2-COL of k-uniform hypergraphs.

Random CSPs

Every clause in a k-SAT formula F (with m clauses on n literals) has the form

$$x_{i_1} \vee x_{i_2} \vee \bar{x}_{i_3} \vee \cdots \vee x_{i_k}.$$

F is satisfiable if there is a solution $(x_1, \dots, x_n)$ such that every clause of F is satisfied.

A random instance F with m clauses and n literals is chosen uniformly at random from the set of $(2n)^{km}$ formulas.

More CSPs:

k-NAESAT, k-XORSAT, k-COL of graphs, 2-COL of k-uniform hypergraphs.

Random CSPs

Every clause in a k-SAT formula F (with m clauses on n literals) has the form

$$x_{i_1} \vee x_{i_2} \vee \bar{x}_{i_3} \vee \cdots \vee x_{i_k}.$$

F is satisfiable if there is a solution $(x_1, \ldots, x_n)$ such that every clause of F is satisfied.

A random instance F with m clauses and n literals is chosen uniformly at random from the set of $(2n)^{km}$ formulas.

More CSPs:

k-NAESAT, k-XORSAT, k-COL of graphs, 2-COL of k-uniform hypergraphs.

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

k-SAT threshold

Theorem (Friedgut 1999)

k-SAT has a sharp threshold at $m/n = c_k(n)$.

- Is it true $\lim_{n\to\infty} c_k(n) = c_k$? (only known: true for $k = 2$)
- What is c_3? (only known: $c_2 = 1$)

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

For large k...

- Kirousis, Kranakis, Krizanc, Stamatiou 1998; Franz, Leone 2003: $c_k \leq 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.
- Achlioptas, Peres 2004: $c_k \geq 2^k \ln 2 - (k/2) \ln 2 - (1 + \ln 2/2) + o_k(1)$.
- Mertens, Mézard, Zecchina 2006: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$ (1RSB).
- Coja-Oghlan, Panagiotou 2013: $c_k \geq 2^k \ln 2 - (3/2) \ln 2 + o_k(1)$.
- Coja-Oghlan 2014: $c_k = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$.

Main challenge: large deviations

- Z: the number of k-SAT solutions;
- $\mathbf{E}Z \to \infty$;
- we need $\sqrt{\mathbf{Var}Z} = O(\mathbf{E}Z)$ to conclude $Z > 0$;
- But $\sqrt{\mathbf{Var}Z} \geq a^n\mathbf{E}Z$.

Main challenge: large deviations

- Z: the number of k-SAT solutions;
- $\mathbf{E}Z \to \infty$;
- we need $\sqrt{\mathbf{Var}Z} = O(\mathbf{E}Z)$ to conclude $Z > 0$;
- But $\sqrt{\mathbf{Var}Z} \geq a^n \mathbf{E}Z$.

Main challenge: large deviations

- Z: the number of k-SAT solutions;
- $\mathbf{E}Z \to \infty$;
- we need $\sqrt{\mathbf{Var}Z} = O(\mathbf{E}Z)$ to conclude $Z > 0$;
- But $\sqrt{\mathbf{Var}Z} \geq a^n \mathbf{E}Z$.

Main challenge: large deviations

- Z: the number of k-SAT solutions;
- $\mathbf{E}Z \to \infty$;
- we need $\sqrt{\mathbf{Var}Z} = O(\mathbf{E}Z)$ to conclude $Z > 0$;
- But $\sqrt{\mathbf{Var}Z} \geq a^n \mathbf{E}Z$.

Main challenge: large deviations

- Z: the number of k-SAT solutions;
- $\mathbf{E}Z \to \infty$;
- we need $\sqrt{\mathbf{Var}Z} = O(\mathbf{E}Z)$ to conclude $Z > 0$;
- But $\sqrt{\mathbf{Var}Z} \geq a^n \mathbf{E}Z$.

What causes large deviations?

- *k*-SAT solutions are not symmetric;
- "Degree sequence";
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

What causes large deviations?

- *k*-SAT solutions are not symmetric;
- "Degree sequence";
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

What causes large deviations?

- k-SAT solutions are not symmetric;
- "Degree sequence";
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

What causes large deviations?

- k-SAT solutions are not symmetric;
- "Degree sequence";
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

What causes large deviations?

- k-SAT solutions are not symmetric;
- “Degree sequence”;
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

What causes large deviations?

- k-SAT solutions are not symmetric;
- "Degree sequence";
- Marginal distribution (The marginal distribution is computed by physicists (cavity method));
- And we need to know more on the solution geometry: **clusters**.

For large k...

	SAT UB	SAT LB	Algorithmic Barrier
k-SAT	$2^k \ln 2 - (1+\ln 2)/2$	$2^k \ln 2 - (1+\ln 2)/2$	$2^k \ln k/k$
k-COL	$2k \ln k - \ln k$	$2k \ln k - \ln k - 2\ln 2$	$(1/2)k \ln k$
2-COL k-hypergraph	$2^{k-1} \ln 2 - \ln 2/2$	$2^{k-1} \ln 2 - (1+\ln 2)/2$	$2^{k-1} \ln k/k$

Connectivity of solutions

Let **x** and **y** be two solutions of F.

The Hamming distance between **x** and **y** is the number of i such that $x_i \neq y_i$.

x	0011101011001010011
y	0101111000001110101

$$d_H(\mathbf{x}, \mathbf{y}) = 8.$$

Connectivity of solutions

Let **x** and **y** be two solutions of F.
The Hamming distance between **x** and **y** is the number of i such that $x_i \neq y_i$.

x	0011101011001010011
y	0101111000001110101

$$d_H(\mathbf{x}, \mathbf{y}) = 8.$$

Connectivity of solutions

Let **x** and **y** be two solutions of F.
The Hamming distance between **x** and **y** is the number of i such that $x_i \neq y_i$.

x 0011101011001010011

y 0101111000001110101

$$d_H(\mathbf{x}, \mathbf{y}) = 8.$$

Connectivity of solutions

Let

- F: a SAT formula.
- $f = f(n)$: any function of n between 1 and n.

Let F_f be the graphs such that

- $V(F_f)$: the set of solutions of F.
- $E(F_f)$: σ and τ are adjacent if $d_H(\sigma, \tau) \leq f$.

If σ and τ are in the same component of F_f, we say they are f-connected.

Connectivity of solutions

Let

- F: a SAT formula.
- $f = f(n)$: any function of n between 1 and n.

Let F_f be the graphs such that

- $V(F_f)$: the set of solutions of F.
- $E(F_f)$: σ and τ are adjacent if $d_H(\sigma, \tau) \leq f$.

If σ and τ are in the same component of F_f, we say they are f-connected.

Connectivity of solutions

Let

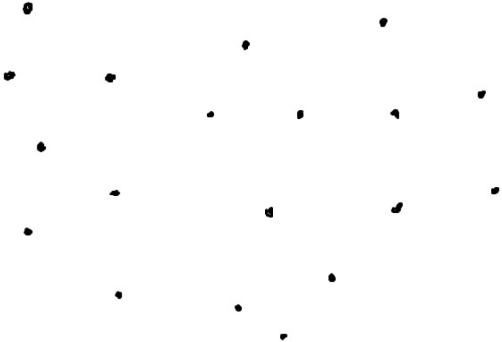
- F: a SAT formula.
- $f = f(n)$: any function of n between 1 and n.

Let F_f be the graphs such that

- $V(F_f)$: the set of solutions of F.
- $E(F_f)$: σ and τ are adjacent if $d_H(\sigma, \tau) \leq f$.

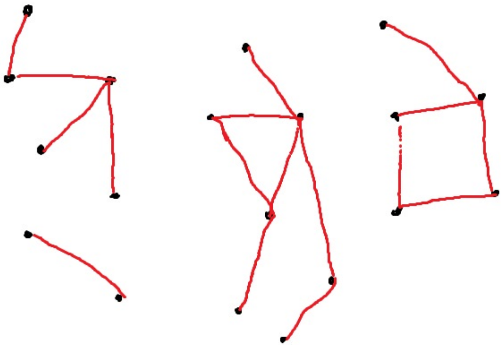
If σ and τ are in the same component of F_f, we say they are f-connected.

Solution connectivity



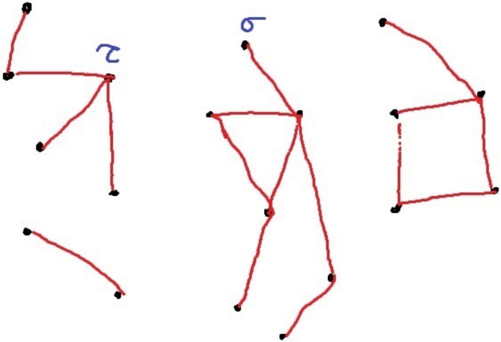
Solution connectivity

F_f



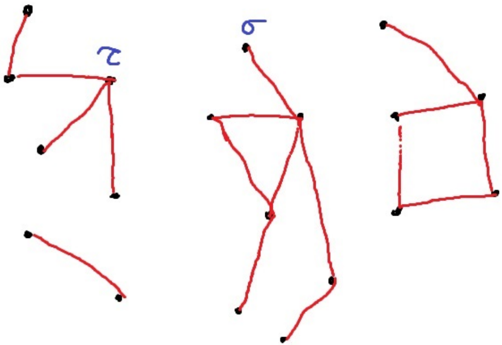
Solution connectivity

F_f $d_H(\tau, \sigma) \geq g$



Solution connectivity

F_f $d_H(\tau, \sigma) \geq g \gg f$



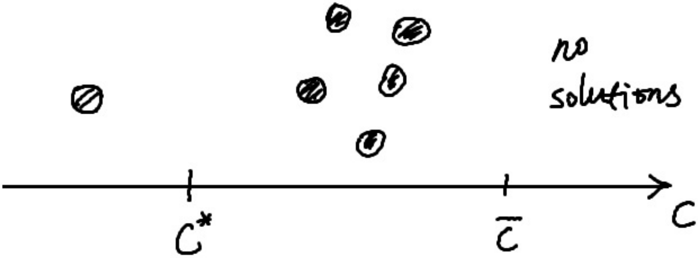
What physicists say...

It was observed and hypothesized (by statistical physicists) that when m/n exceeds a certain threshold, the solution space of many CSPs is partitioned into clusters.

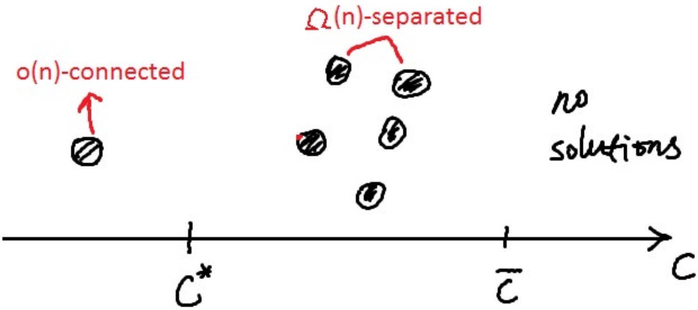
What physicists say...

It was observed and hypothesized (by statistical physicists) that when m/n exceeds a certain threshold, the solution space of many CSPs is partitioned into clusters.

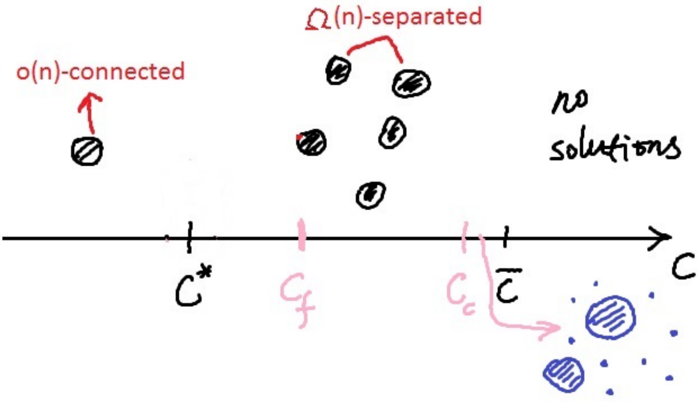
Clustering picture



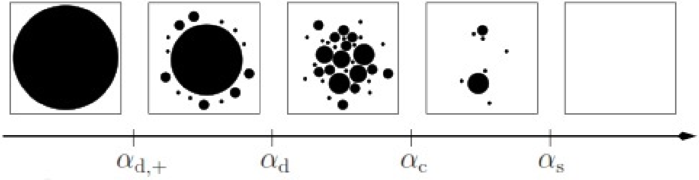
Clustering picture



Clustering picture



Clustering picture

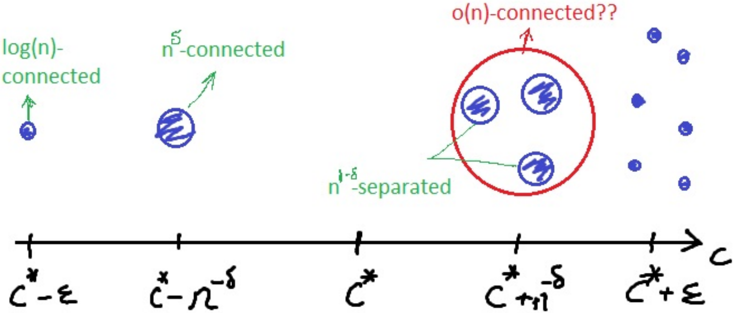


[Krzakala, Montanari, Ricci-Tersenghi, Semerjian, Zdeborova 2007]

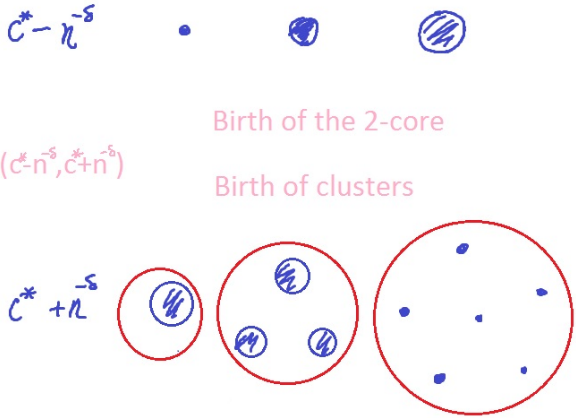
What mathematicians say...

- k-XORSAT clustering threshold determined (AM12,IKKM12)
- k-SAT has well-separated clusters (DMMZ08,MMZ05)
- k-SAT (NAESAT,COL) clusters appear after (asymptotically in k) the hypothesized clustering threshold (AC08)
- k-SAT clusters contains frozen variables (AR09)
- freezing threshold for k-COL, k-NAESAT (MR13)
- Condensation occurs in 2-COL in hypergraph (CZ12)

Birth of *k*-XORSAT clusters



Birth of k-XORSAT clusters: $c^* - \epsilon < c < c^* + \epsilon$

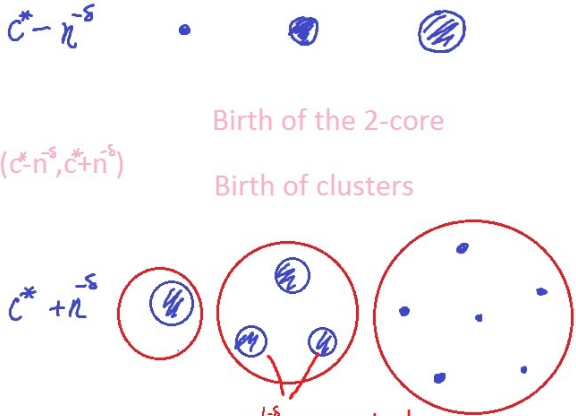


Birth of the 2-core

$(c^*-n^{-\delta}, c^*+n^{-\delta})$

Birth of clusters

Birth of k-XORSAT clusters: $c^* - \epsilon < c < c^* + \epsilon$



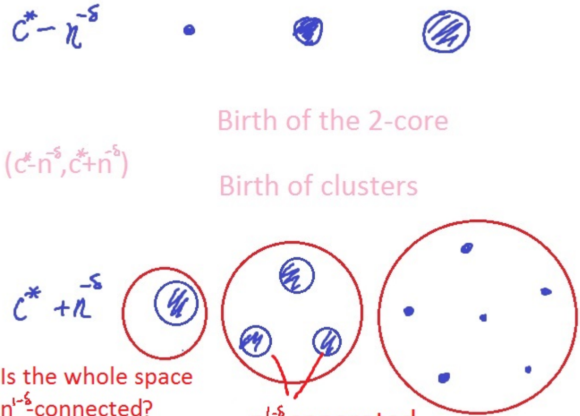
Birth of the 2-core

$(c^* - n^{-\delta}, c^* + n^{-\delta})$

Birth of clusters

$n^{1-\delta}$-separated

Birth of k-XORSAT clusters: $c^* - \epsilon < c < c^* + \epsilon$



Birth of the 2-core

$(c^* - n^{-\delta}, c^* + n^{-\delta})$

Birth of clusters

Is the whole space $n^{1-\delta}$-connected?

$n^{1-\delta}$-separated

Linear system vs XORSAT formula

A clause in an k-XORSAT formula F (with m clauses) has the form

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k,$$

where $x_1 \oplus x_2$ is true $(= 1)$ if and only if exactly one of the variables is true $(= 1)$.
F is linear:

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k = x_1 + x_2 + (1 + x_3) + \cdots + x_k (mod\ 2).$$

An assignment satisfying F is a solution to a linear system of the following form

$$A\mathbf{x} = \mathbf{b} (\text{mod } 2),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of boolean variables;
A is an $m \times n$ matrix.

Linear system vs XORSAT formula

A clause in an k-XORSAT formula F (with m clauses) has the form

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k,$$

where $x_1 \oplus x_2$ is true ($= 1$) if and only if exactly one of the variables is true ($= 1$).

F is linear:

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k = x_1 + x_2 + (1 + x_3) + \cdots + x_k (mod\ 2).$$

An assignment satisfying F is a solution to a linear system of the following form

$$A\mathbf{x} = \mathbf{b} (\mathrm{mod}\ 2),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of boolean variables; A is an $m \times n$ matrix.

Linear system vs XORSAT formula

A clause in an k-XORSAT formula F (with m clauses) has the form

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k,$$

where $x_1 \oplus x_2$ is true $(= 1)$ if and only if exactly one of the variables is true $(= 1)$.
F is linear:

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k = x_1 + x_2 + (1 + x_3) + \cdots + x_k (mod\ 2).$$

An assignment satisfying F is a solution to a linear system of the following form

$$A\mathbf{x} = \mathbf{b} (\text{mod } 2),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of boolean variables;
A is an $m \times n$ matrix.

Linear system vs XORSAT formula

A clause in an k-XORSAT formula F (with m clauses) has the form

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k,$$

where $x_1 \oplus x_2$ is true $(= 1)$ if and only if exactly one of the variables is true $(= 1)$.
F is linear:

$$x_1 \oplus x_2 \oplus \bar{x}_3 \oplus \cdots \oplus x_k = x_1 + x_2 + (1 + x_3) + \cdots + x_k (mod\ 2).$$

An assignment satisfying F is a solution to a linear system of the following form

$$A\mathbf{x} = \mathbf{b} (\text{mod } 2),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of boolean variables;
A is an $m \times n$ matrix.

Linear system vs hypergraph

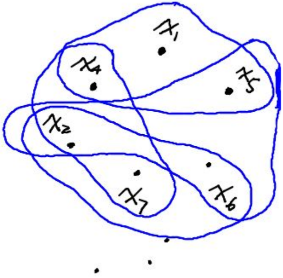
$$x_1 + x_4 + x_5 = 1$$

$$x_2 + x_4 + x_7 = 0$$

$$x_2 + x_7 + x_6 = 0$$

$$x_5 + x_6 + x_2 = 1$$

$$\vdots$$



Linear system vs hypergraph

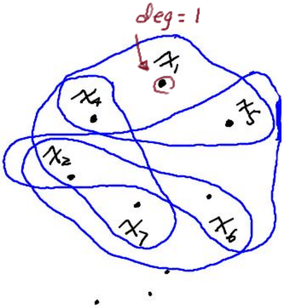
$\boxed{x_1} + x_4 + x_5 = 1$

$x_2 + x_4 + x_7 = 0$

$x_2 + x_7 + x_6 = 0$

$x_5 + x_6 + x_2 = 1$

$\vdots$



deg = 1

Linear system vs hypergraph

$$x_1 = 1 - x_4 - x_5$$

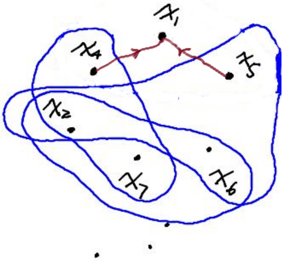
~~$\boxed{x_1} + x_4 + x_5 = 1$~~

$$x_2 + x_4 + x_7 = 0$$

$$x_2 + x_7 + x_6 = 0$$

$$x_5 + x_6 + x_2 = 1$$

$$\vdots$$



Linear system vs hypergraph

$x_1 = 1 - x_4 - x_5$

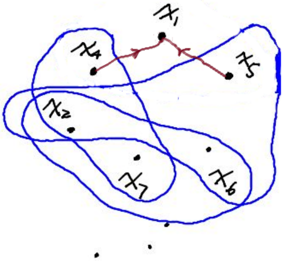
~~$\boxed{x_1} + x_4 + x_5 = 1$~~

$x_2 + \boxed{x_4} + x_7 = 0$

$x_2 + x_7 + x_6 = 0$

$x_5 + x_6 + x_2 = 1$

$\vdots$



Linear system vs hypergraph

$x_1 = 1 - x_4 - x_5$

$x_4 = - x_2 - x_7$

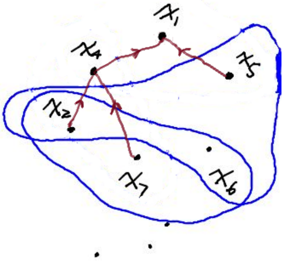
~~$\boxed{x_1} + x_4 + x_5 = 1$~~

~~$x_2 + \boxed{x_4} + x_7 = 0$~~

$x_2 + x_7 + x_6 = 0$

$x_5 + x_6 + x_2 = 1$

⋮



Linear system vs hypergraph

$x_1 = 1 - x_4 - x_5$

$x_4 = - x_2 - x_7$

~~$\boxed{x_1} + x_4 + x_5 = 1$~~

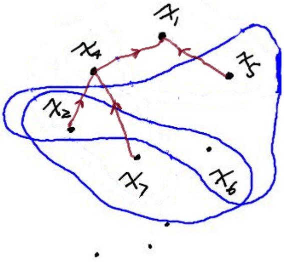
~~$x_2 + \boxed{x_4} + x_7 = 0$~~

$x_2 + x_7 + x_6 = 0$

$x_5 + x_6 + x_2 = 1$

$\vdots$

$A' x = b'$



2-core of H

Clusters

k-XORSAT clusters: the set of solutions extended from a particular solution on the 2-core (plus some complications).

Clusters

k-XORSAT clusters: the set of solutions extended from a particular solution on the 2-core (plus some complications).

The following is proved in dependently by Achlioptas and Molloy, and Ibrahimi et al. about random *r*-XORSAT clustering.

Theorem (AM12,IKKM12)

A.a.s. the following statements are true.

(a) *If $m < (c^* - \epsilon)n$, then all solutions are in a single $O(\log n)$-connected cluster.*

(b) *If $m > (c^* + \epsilon)n$, then all solutions are partitioned into well-connected well-separated clusters.(all clusters are $O(\log n)$-connected; every pair of clusters are $\Omega(n)$-separated)*

Clusters

k-XORSAT clusters: the set of solutions extended from a particular solution on the 2-core (plus some complications). The following is proved in dependently by Achlioptas and Molloy, and Ibrahimi et al. about random r-XORSAT clustering.

Theorem (AM12,IKKM12)

A.a.s. the following statements are true.

(a) *If* $m < (c^* - \epsilon)n$*, then all solutions are in a single* $O(\log n)$*-connected cluster.*

(b) *If* $m > (c^* + \epsilon)n$*, then all solutions are partitioned into well-connected well-separated clusters.(all clusters are* $O(\log n)$*-connected; every pair of clusters are* $\Omega(n)$*-separated)*

Clusters

k-XORSAT clusters: the set of solutions extended from a particular solution on the 2-core (plus some complications).
The following is proved in dependently by Achlioptas and Molloy, and Ibrahimi et al. about random r-XORSAT clustering.

Theorem (AM12,IKKM12)

A.a.s. the following statements are true.

(a) *If $m < (c^* - \epsilon)n$, then all solutions are in a single $O(\log n)$-connected cluster.*

(b) *If $m > (c^* + \epsilon)n$, then all solutions are partitioned into well-connected well-separated clusters.(all clusters are $O(\log n)$-connected; every pair of clusters are $\Omega(n)$-separated)*

c^* corresponds to the emergence threshold of the 2-core.

2-core threshold

Theorem (Kim '06)

Let H be a random k-uniform hypergraph with n vertices and cn edges. There is a constant $c^* > 0$ (depending on k and can be specified) such that

(a) *if $c < c^* - n^{-1/2+\epsilon}$, then w.h.p. H has an empty 2-core;*

(b) *if $c > c^* + n^{-1/2+\epsilon}$, then w.h.p. H has a 2-core with size $\Omega(n)$.*

Clusters

Theorem (**G.** and Molloy '13)

Suppose $m = (c^* + n^{-\delta})n$, $\delta > 0$ *small. Then, a.a.s.*

(a) *every cluster is* $n^{O(\delta)}$*-connected;*

(b) *every pair of clusters are* $n^{1-O(\delta)}$*-separated;*

(c) *(a) is tight – in every cluster, there are two solutions that are not* $n^{\Theta(\delta)}$*-connected.*

Clusters

Theorem (**G.** and Molloy '13)

Suppose $m = (c^* + n^{-\delta})n$, $\delta > 0$ *small. Then, a.a.s.*

(a) *every cluster is* $n^{O(\delta)}$*-connected;*

(b) *every pair of clusters are* $n^{1-O(\delta)}$*-separated;*

(c) *(a) is tight – in every cluster, there are two solutions that are not* $n^{\Theta(\delta)}$*-connected.*

Inside the cluster-Upper bound

$x_1 = 1 - x_4 - x_5$

$x_4 = - x_2 - x_7$

~~$\boxed{x_1} + x_4 + x_5 = 1$~~

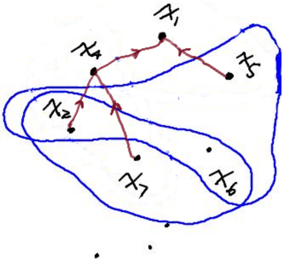
~~$x_2 + \boxed{x_4} + x_7 = 0$~~

$x_2 + x_7 + x_6 = 0$

$x_5 + x_6 + x_2 = 1$

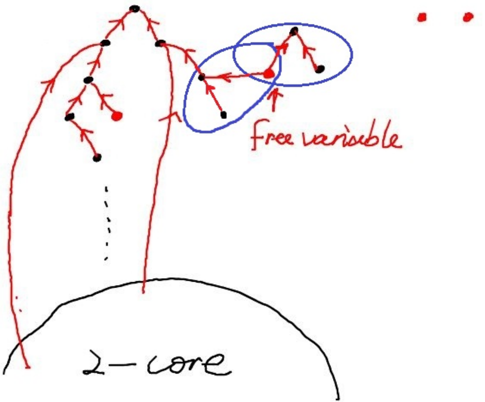
$\vdots$

$A'x = b'$

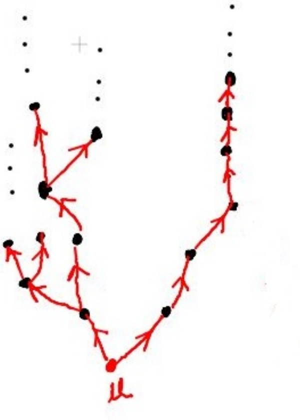


2-core of H

Inside the cluster-Upper bound



Inside the cluster-Upper bound



Inside the cluster-Upper bound

$depth(v)$: the least number of vertices needs to be removed before v in any stripping sequence. (These are the set of vertices reachable from v in the digraph.)

When $c = c^* + \epsilon$, $depth(v) = O(\log n)$ for all $v \notin$ 2-core;

When $c = c^* + n^{-\delta}$, the maximum depth becomes $n^{\Theta(\delta)}$.

Inside the cluster-Upper bound

depth(v): the least number of vertices needs to be removed before v in any stripping sequence. (These are the set of vertices reachable from v in the digraph.)
When $c = c^* + \epsilon$, $depth(v) = O(\log n)$ for all $v \notin$ 2-core;
When $c = c^* + n^{-\delta}$, the maximum depth becomes $n^{\Theta(\delta)}$.

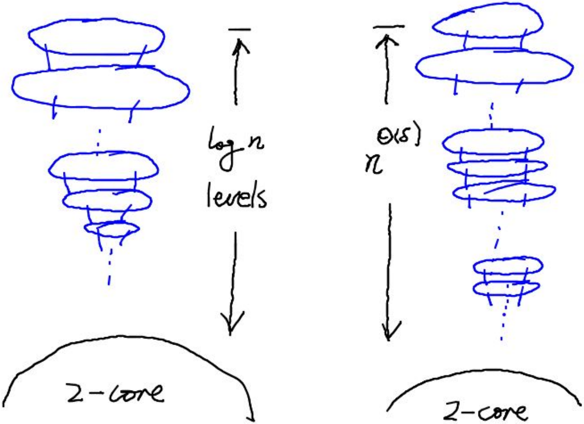
Inside the cluster-Upper bound

$depth(v)$: the least number of vertices needs to be removed before v in any stripping sequence. (These are the set of vertices reachable from v in the digraph.)
When $c = c^* + \epsilon$, $depth(v) = O(\log n)$ for all $v \notin$ 2-core;
When $c = c^* + n^{-\delta}$, the maximum depth becomes $n^{\Theta(\delta)}$.

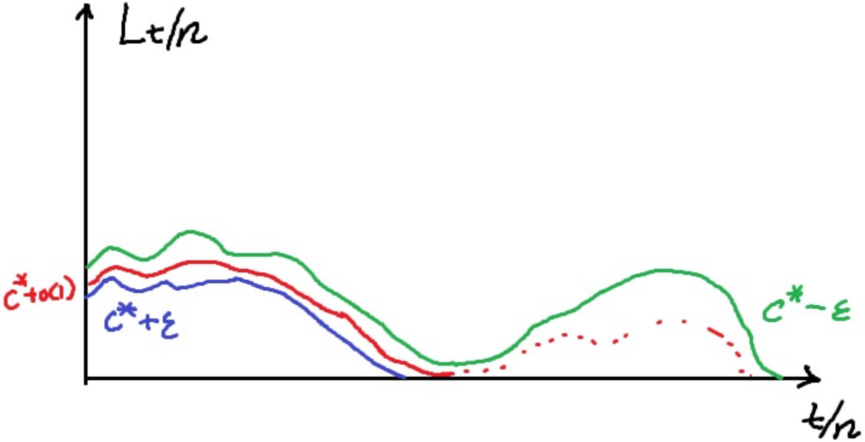
Inside the cluster-Upper bound

depth(v): the least number of vertices needs to be removed before v in any stripping sequence. (These are the set of vertices reachable from v in the digraph.)
When $c = c^* + \epsilon$, *depth*$(v) = O(\log n)$ for all $v \notin$ 2-core;
When $c = c^* + n^{-\delta}$, the maximum depth becomes $n^{\Theta(\delta)}$.

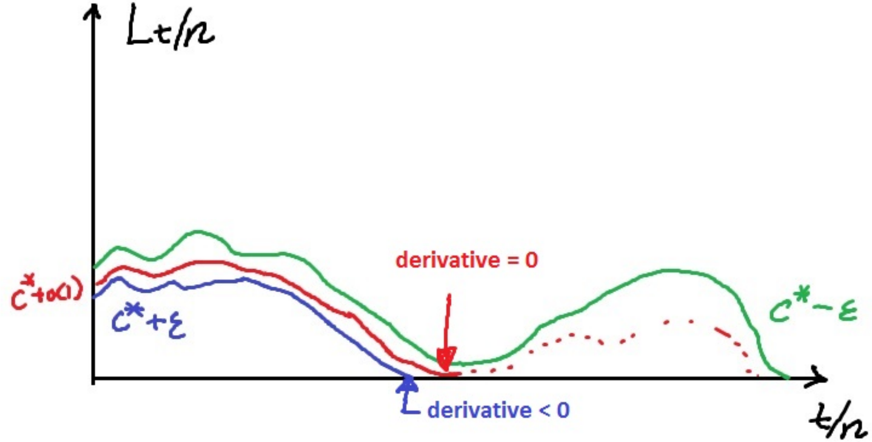
Lower bound of maximum depth



Why stripping becomes slow?



Why stripping becomes slow?



Stripping number

Theorem (G. and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $0 < \delta < 1/2$, *the stripping number is between* $n^{\delta/2}$ *and* $n^{\delta/2} \log n$.

Idea behind the proof Approximate L_t by a "lazy" random walk – it takes long for L_t to reach 0. ☐

It implies the lower bound on the maximum depth.

Stripping number

Theorem (G. and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $0 < \delta < 1/2$, the stripping number is between $n^{\delta/2}$ and $n^{\delta/2} \log n$.

Idea behind the proof Approximate L_t by a "lazy" random walk – it takes long for L_t to reach 0. ☐

It implies the lower bound on the maximum depth.

Stripping number

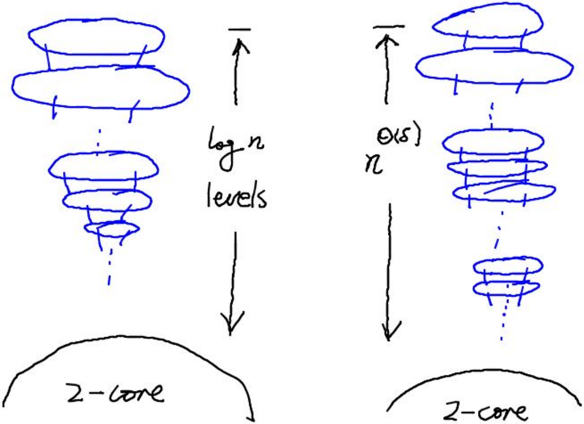
Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $0 < \delta < 1/2$, *the stripping number is between* $n^{\delta/2}$ *and* $n^{\delta/2} \log n$*.*

Idea behind the proof Approximate L_t by a "lazy" random walk – it takes long for L_t to reach 0. ☐

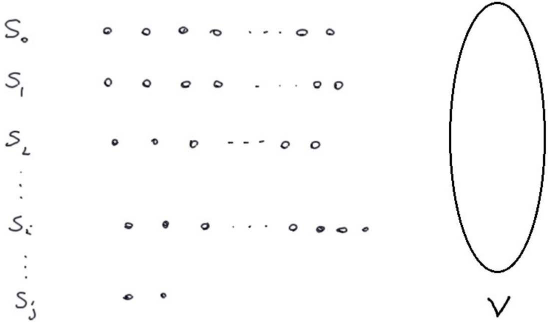
It implies the lower bound on the maximum depth.

Upper bound of maximum depth



Upper bound of maximum depth

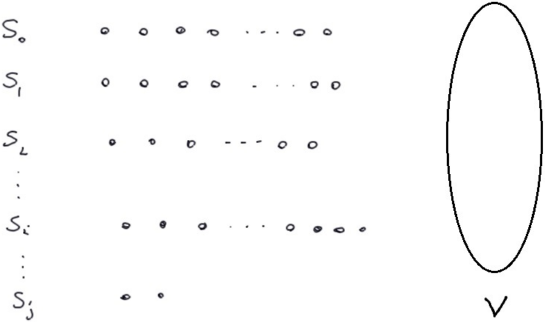
Expose $S_0, S_1, \cdots S_j, V$



Upper bound of maximum depth

Expose S_0, S_1, $\cdots$ S_j. V

Expose $d_i(v)$, $d^+(v)$, $d^{++}(v)$

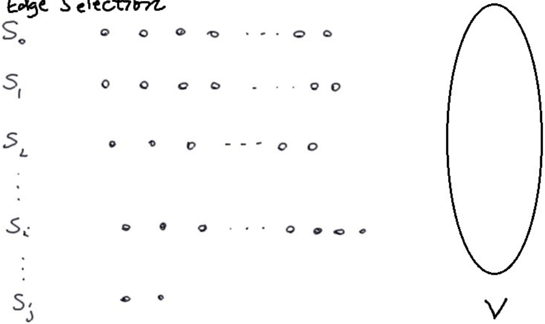


Upper bound of maximum depth

Expose $S_0, S_1, \cdots S_j, V$

Expose $d_i(v)$, $d^+(v)$, $d^{++}(v)$

Edge Selection

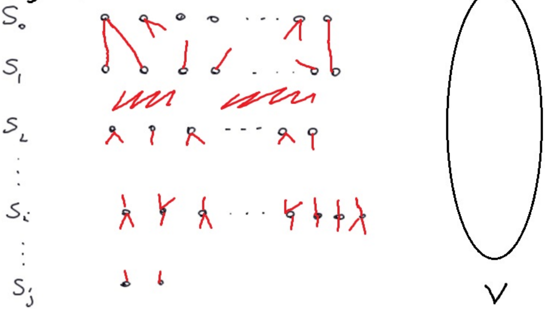


Upper bound of maximum depth

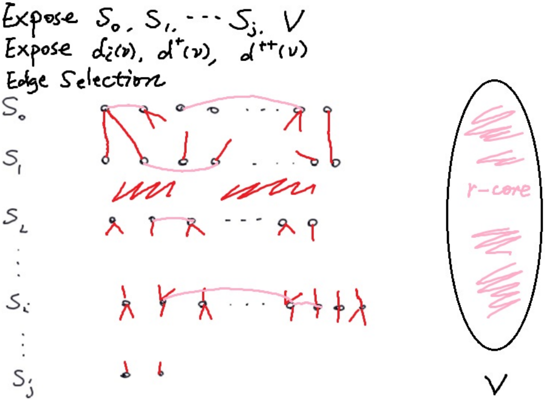
Expose S_0, S_1, $\cdots$ S_j, V

Expose $d_i(v)$, $d^+(v)$, $d^{++}(v)$

Edge Selection



Upper bound of maximum depth



Expose $S_0, S_1, \cdots S_j, V$

Expose $d_i(v)$, $d^+(v)$, $d^{++}(v)$

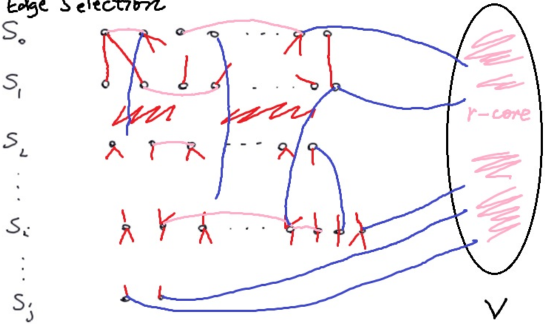
Edge Selection

Upper bound of maximum depth

Expose S_0. S_1. $\cdots$ S_j. V

Expose $d_i(v)$, $d^{+}(v)$, $d^{++}(v)$

Edge Selection



Upper bound of maximum depth

Theorem (G. and Molloy '13)

Suppose $c = c^* + n^{-\delta}$, $\delta > 0$ *small.*

(a) $depth(v) = n^{O(\delta)}$ *for all* $v \notin$ *2-core;*

(b) $\max\{depth(v)\} = n^{\Theta(\delta)}$.

Proof (a). Use Edge-Selection + solving a recurrence + applying Talagrand Inequality.
(b). From the lower bound of the stripping number. ☐

Upper bound of maximum depth

Theorem (**G.** and Molloy '13)

Suppose $c = c^* + n^{-\delta}$, $\delta > 0$ *small.*

(a) $depth(v) = n^{O(\delta)}$ *for all* $v \notin$ *2-core;*

(b) $\max\{depth(v)\} = n^{\Theta(\delta)}$.

Proof (a). Use Edge-Selection + solving a recurrence + applying Talagrand Inequality.

(b). From the lower bound of the stripping number. ☐

Upper bound of maximum depth

Theorem (**G.** and Molloy '13)

Suppose $c = c^* + n^{-\delta}$, $\delta > 0$ *small.*

(a) $depth(v) = n^{O(\delta)}$ *for all* $v \notin$ *2-core;*

(b) $\max\{depth(v)\} = n^{\Theta(\delta)}$.

Proof (a). Use Edge-Selection + solving a recurrence + applying Talagrand Inequality.
(b). From the lower bound of the stripping number. ☐

Inside the cluster

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $\delta > 0$ *small. Then a.a.s.*

(a) *every cluster is* $n^{O(\delta)}$*-connected;*

(b) *there exists two solutions in each cluster that are* **NOT** $n^{\Theta(\delta)}$*-connected.*

Proof (a) follows as the maximum depth of vertices outside the 2-core is $n^{O(\delta)}$. □

However, the lower bound of the maximum depth does **NOT** imply part (b).

Inside the cluster

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $\delta > 0$ *small. Then a.a.s.*

(a) *every cluster is* $n^{O(\delta)}$*-connected;*

(b) *there exists two solutions in each cluster that are* **NOT** $n^{\Theta(\delta)}$*-connected.*

Proof (a) follows as the maximum depth of vertices outside the 2-core is $n^{O(\delta)}$. ☐

However, the lower bound of the maximum depth does **NOT** imply part (b).

Inside the cluster

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $\delta > 0$ *small. Then a.a.s.*

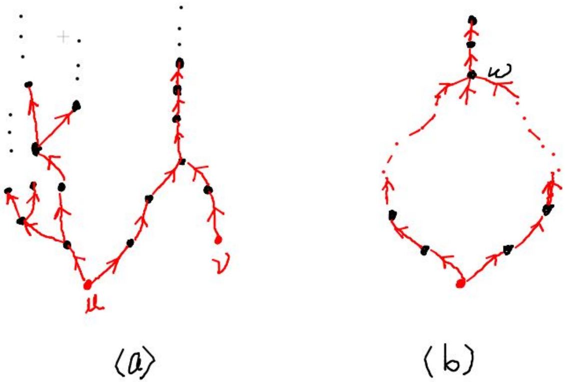
(a) *every cluster is* $n^{O(\delta)}$*-connected;*

(b) *there exists two solutions in each cluster that are* **NOT** $n^{\Theta(\delta)}$*-connected.*

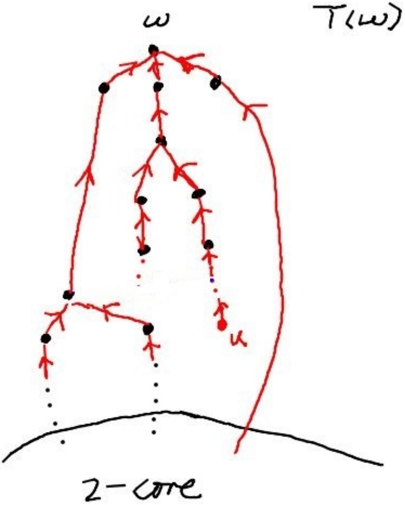
Proof (a) follows as the maximum depth of vertices outside the 2-core is $n^{O(\delta)}$. □

However, the lower bound of the maximum depth does **NOT** imply part (b).

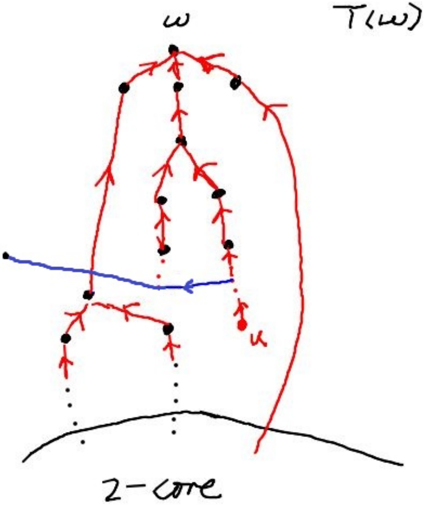
Here is why?



The way to cope with it...



The way to cope with it...



Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ small. Then a.a.s. each pair of clusters is $n^{1-O(\delta)}$-separated.

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

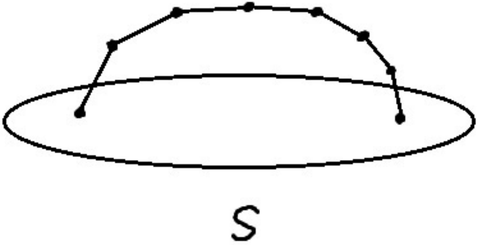
Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

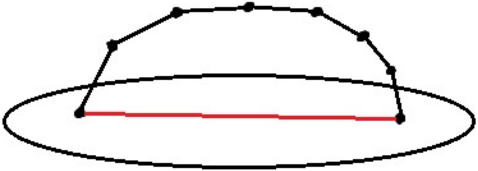
$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

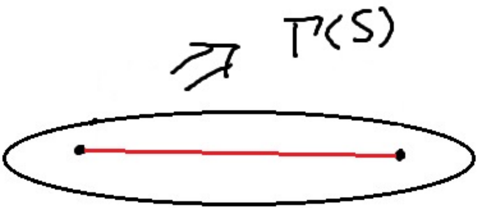
Proof



Proof



Proof



Between clusters

Theorem (G. and Molloy 2013)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

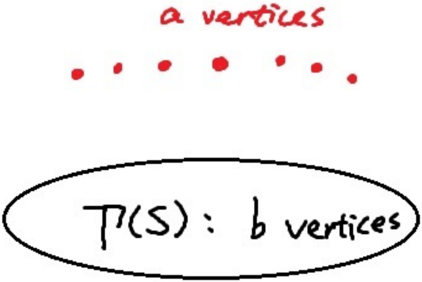
Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;
$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;
$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;
$\Rightarrow$ $\Gamma(S)$;
$\Rightarrow$ average degree of $\Gamma(S)$ is at least 3;

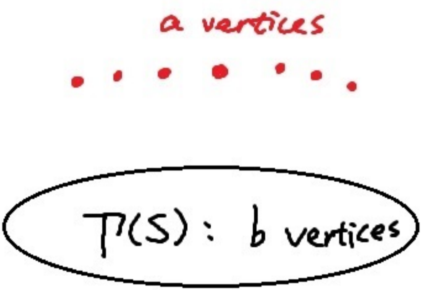
Proof

Proof



case 1: a=O(b), both a, b small.

Proof



case 1: a=O(b), both a, b small.

⇒ a small graph: average degree >2

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

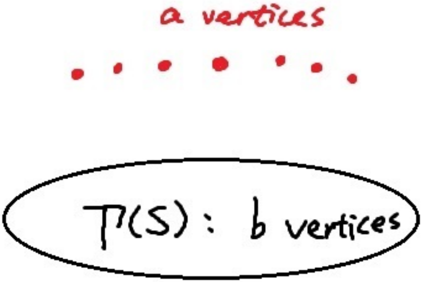
$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

$\Rightarrow$ average degree of $\Gamma(S)$ is at least 3;

(a) A sparse random graph is locally tree-like;

Proof

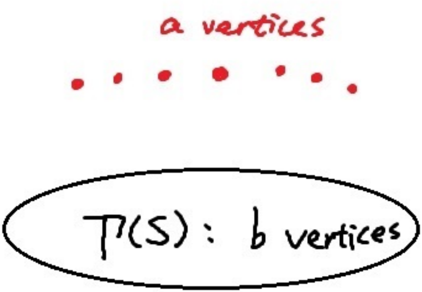


case 1: a=O(b), both a, b small.

⇒ a small graph: average degree >2

case 2: b small, a>> b.

Proof



case 1: a=O(b), both a, b small.

⇒ a small graph: average degree >2

case 2: b small, a>> b.

Between clusters

Theorem (**G.** and Molloy '13)

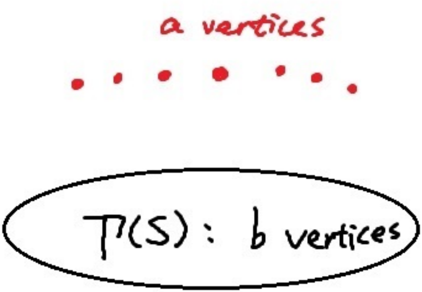
Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;
$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;
$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;
$\Rightarrow$ $\Gamma(S)$;
$\Rightarrow$ average degree of $\Gamma(S)$ is at least 3;
(a) A sparse random graph is locally tree-like; (b) Branching parameter $= 1 - \zeta$;

Proof



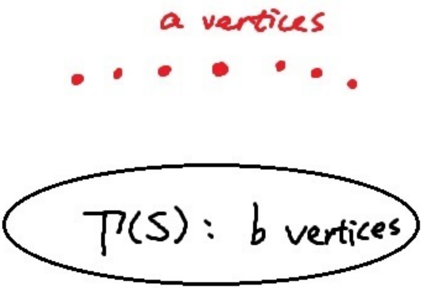
case 1: $a=O(b)$, both a, b small.

⇒ a small graph: average degree >2

case 2: b small, a>> b.

⇒ such a path cannot be long

Proof



case 1: a=O(b), both a, b small.

⇒ a small graph: average degree >2

case 2: b small, a>> b.

⇒ such a path cannot be long

cases 1,2 ⇒ a+b must be large

Between clusters

Theorem (**G.** and Molloy '13)

Assume $c = c^* + n^{-\delta}$, $r \geq 3$, $\delta > 0$ *small. Then a.a.s. each pair of clusters is* $n^{1-O(\delta)}$*-separated.*

Note that this theorem does not exclude the possibility that clusters are $\Omega(n)$-separated.

Proof

Take x and y disagree on the 2-core;

$\Rightarrow$ x and y disagree on a set S of variables in the 2-core;

$\Rightarrow$ $|e \cap S|$ is even for every e in the 2-core;

$\Rightarrow$ $\Gamma(S)$;

$\Rightarrow$ average degree of $\Gamma(S)$ is at least 3;

(a) A sparse random graph is locally tree-like; (b) Branching parameter $= 1 - \zeta$;

$\Rightarrow$ x and y are well-separated.

Future work

- For $c = c^* + n^{-\delta}$, are clusters $o(n)$-connected or $\Omega(n)$-separated?
- $c = c^* - n^{-\delta}$ (stripping number, depth, cluster connectivity)?
- $c = c^* + O(n^{-1/2})$ (stripping number, depth, cluster connectivity)?
- What happens to the first graph containing a 2-core?

Future work

- For $c = c^* + n^{-\delta}$, are clusters $o(n)$-connected or $\Omega(n)$-separated?
- $c = c^* - n^{-\delta}$ (stripping number, depth, cluster connectivity)?
- $c = c^* + O(n^{-1/2})$ (stripping number, depth, cluster connectivity)?
- What happens to the first graph containing a 2-core?

Future work

- For $c = c^* + n^{-\delta}$, are clusters $o(n)$-connected or $\Omega(n)$-separated?
- $c = c^* - n^{-\delta}$ (stripping number, depth, cluster connectivity)?
- $c = c^* + O(n^{-1/2})$ (stripping number, depth, cluster connectivity)?
- What happens to the first graph containing a 2-core?

Future work

- For $c = c^* + n^{-\delta}$, are clusters $o(n)$-connected or $\Omega(n)$-separated?
- $c = c^* - n^{-\delta}$ (stripping number, depth, cluster connectivity)?
- $c = c^* + O(n^{-1/2})$ (stripping number, depth, cluster connectivity)?
- What happens to the first graph containing a 2-core?

Future work

- For $c = c^* + n^{-\delta}$, are clusters $o(n)$-connected or $\Omega(n)$-separated?
- $c = c^* - n^{-\delta}$ (stripping number, depth, cluster connectivity)?
- $c = c^* + O(n^{-1/2})$ (stripping number, depth, cluster connectivity)?
- What happens to the first graph containing a 2-core?