Mathematical problems of very large networks

László Lovász

Eötvös Loránd University, Budapest

lovasz@cs.elte.hu

Issues on very large graphs

The following issues are closely related:

- property testing;
- parameter estimation;
- limit objects for convergent graph sequences;
- regularity lemmas;
- distance of graphs;
- duality of left and right convergence.

Issues on very large graphs

The following concepts are cryptomorphic:

- a consistent local finite random graph model;
- a consistent local countable random graph;
- a measurable, symmetric function W: $[0,1]^2 \to [0,1]$;
- a multiplicative graph parameter with nonnegative Möbius transform;
- a multiplicative, reflection positive graph parameter;
- A point in the completion of the set of finite graphs with the cut-distance.

Cut distance of two graphs

(a) $V(G) = V(G')$

$$d_{\square}(G,G') = \max_{S,T\subseteq V(G)} \frac{|e_G(S,T) - e_{G'}(S,T)|}{n^2}$$

(b) $|V(G)| = |V(G')| = n$

$$\delta_{\square}^{*}(G,G') = \min_{G\leftrightarrow G'} d_{\square}(G,G')$$

Cut distance of two graphs

(c) $|V(G)| = n \neq n' = |V(G')|$

blow up nodes, or fractional overlay

$$(X_{ij})_{i\in V(G),\, j\in V(G')} \geq 0 \qquad \sum_{i\in V(G)} X_{ij} = \frac{1}{n'}, \quad \sum_{j\in V(G')} X_{ij} = \frac{1}{n}$$

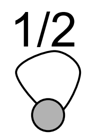
$$\delta_\square(G,G') =$$

$$= \min_X \max_{S,T\subseteq V(G)\times V(G')} \left| \sum_{(i,u)\in S} \sum_{(j,v)\in T} X_{iu} X_{jv} (a_{ij} - a'_{uv}) \right|$$

Cut distance of two graphs

Examples: $\delta_\square(K_{n,n}, \mathbb{G}(2n, \tfrac{1}{2})) \approx \dfrac{1}{8}$

$$\delta_\square\left(\mathbb{G}_1(n, \tfrac{1}{2}), \mathbb{G}_2(n, \tfrac{1}{2})\right) = o(1)$$

$\delta_\square\left(\mathbb{G}_1(n, \tfrac{1}{2}),\right.$ 

$\left.\right) = \delta_\square\left(\mathbb{G}_1(n, \tfrac{1}{2}), 1/2\right) = o(1)$

Sampling Lemmas

G, G': graphs with $V(G) = V(G')$

$\mathbf{S}_k \subseteq V(G)$: random set of k nodes

With large probability,

$$\left| d_\square(G[\mathbf{S}_k], G'[\mathbf{S}_k]) - d_\square(G, G') \right| < \frac{10}{k^{1/4}}$$

Alon-Fernandez de la Vega-Kannan-Karpinski+

With large probability,

$$\delta_\square(G, G[\mathbf{S}_k]) < \frac{10}{\sqrt{\log k}}$$

Borgs-Chayes-Lovász-Sós-Vesztergombi

Regularity Lemmas

Original Regularity Lemma	Szemerédi 1976
"Weak" Regularity Lemma	Frieze-Kannan 1999
"Strong" Regularity Lemma	Alon – Fischer - Krivelevich - M. Szegedy 2000

Regularity Lemmas

$\mathcal{P} = \{V_1, ..., V_k\}$ partition of $V(G)$:

$G_{\mathcal{P}}$ is the complete graph on $V(G)$ with edgeweights

$$w_{uv} = \frac{e_G(V_i, V_j)}{|V_i| \cdot |V_j|} \quad (u \in V_i, v \in V_j)$$

Regularity Lemmas

“Weak" Regularity Lemma (Frieze-Kannan):

For every graph G and $k \geq 1$
there is a partition $\mathcal{P}$ of $V(G)$ with k classes such that

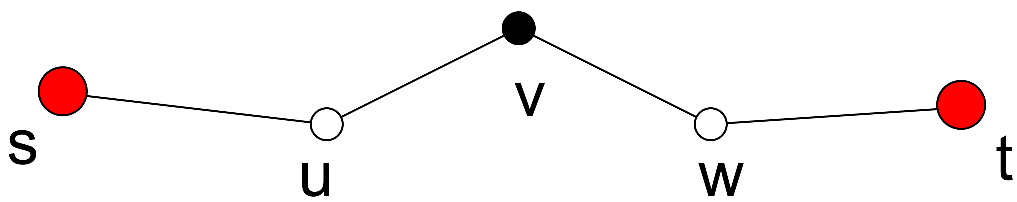
$$\delta_\square(G, G_\mathcal{P}) \leq \frac{1}{\sqrt{\log k}}.$$

For every graph G and $k \geq 1$
there is a graph H with k nodes such that

$$\delta_\square(G, H) \leq \frac{2}{\sqrt{\log k}}.$$

"Weak" Regularity Lemma: geometric form

$$d_2(s,t) := \mathsf{E}_v \left| \mathsf{E}_u (a_{su} a_{vu}) - \mathsf{E}_w (a_{tw} a_{wv}) \right|$$



Fact 1. This is a metric, computable by sampling

Fact 2.
Weak Szemerédi partition $\leftrightarrow$ partition most nodes into sets with small diameter

"Weak" Regularity Lemma: geometric form

$S \subseteq [0,1]$: $\quad \bar{d}(S) = \mathsf{E}_x d(x,S)$ — average ε-net

$\mathcal{P}$ partition of $[0,1]$: $\quad r(\mathcal{P}) = d_\square(G, G_\mathcal{P})$ — regular partition

$\forall S \subseteq [0,1] \Rightarrow$ Voronoi cells of S form a partition with

$$r(\mathcal{P}) < 8\sqrt{\bar{d}(S)}$$

$\forall$ partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of $[0,1]$ $\exists\, v_i \in V_i$ with

$$\bar{d}(\{v_1, \ldots, v_k\}) < 12 r(\mathcal{P})$$

LL – B. Szegedy

"Weak" Regularity Lemma: algorithm

Algorithm to construct representatives of classes:

- Begin with $U=\emptyset$.
- Select random nodes $v_1, v_2, \ldots$
- Put v_i in U iff $d_2(v_i,u)>\varepsilon$ for all $u\in U$.
- Stop if for more than $1/\varepsilon^2$ trials, U did not grow.

size bounded by O(min # classes)

“Weak” Regularity Lemma: algorithm

Algorithm to decide in which class *v* belongs:

Let $U=\{u_1,\ldots,u_k\}$.

Put a node v in V_i iff u_i is the nearest node to *v* in *U*.

Max Cut in huge graphs

(Different algorithm implicit by Frieze-Kannan.)

Algorithm to construct representation of cut:

- Construct U as for the weak Szemerédi partition

- Compute p_{ij} = density between classes V_i and V_j (use sampling)

- Compute max cut (U_1,U_2) in complete graph on U with edge-weights p_{ij}

Max Cut in huge graphs

Algorithm to decide in which class does v belong:

- Put $v \in V$ into V_1 if $d_2(v,U_1) \le d_2(v,U_2)$

  V_2 if $d_2(v,U_1) > d_2(v,U_2)$

Convergent graph sequences

$\hom(G,H) :=$ # of homomorphisms of G into H

$$t(F,G) = \frac{\hom(F,G)}{|V(G)|^{|V(F)|}}$$

Probability that random map $V(F) \to V(G)$ is a hom

(i) $(G_1, G_2, \ldots)$ convergent: Cauchy in the $\delta_\square$-metric.

(ii) $(G_1, G_2, \ldots)$ convergent: $\forall F \;\; t(F, G_n)$ is convergent

distribution of *k*-samples is convergent for all *k*

(i) and (ii) are equivalent.

Convergent graph sequences

Example: random graphs

$$t(F, \mathbb{G}(n, \tfrac{1}{2})) \to \left(\tfrac{1}{2}\right)^{|E(F)|}$$ with probability 1

$$\delta_{\square}(\mathbb{G}(n, \tfrac{1}{2}), \mathbb{G}(m, \tfrac{1}{2})) \to 0 \qquad (n, m \to \infty)$$

Convergent graph sequences

(i) and (ii) are equivalent.

"Counting lemma": $|t(F,G)-t(F,H)| \le |E(F)|\,\delta_\square(G,H)$

"Inverse counting lemma": if $|t(F,G)-t(F,H)| \le \frac{1}{k}$

for all graphs *F* with *k* nodes, then $\delta_\square(G,H) < \frac{10}{\sqrt{\log k}}$

Limit objects

- a consistent local finite random graph model

$G[\mathbf{S}_k]$: probability distribution on k-point graphs

consistent

(a) $G[\mathbf{S}_k] \setminus \{v\}$ has same distribution as $G[\mathbf{S}_{k-1}]$

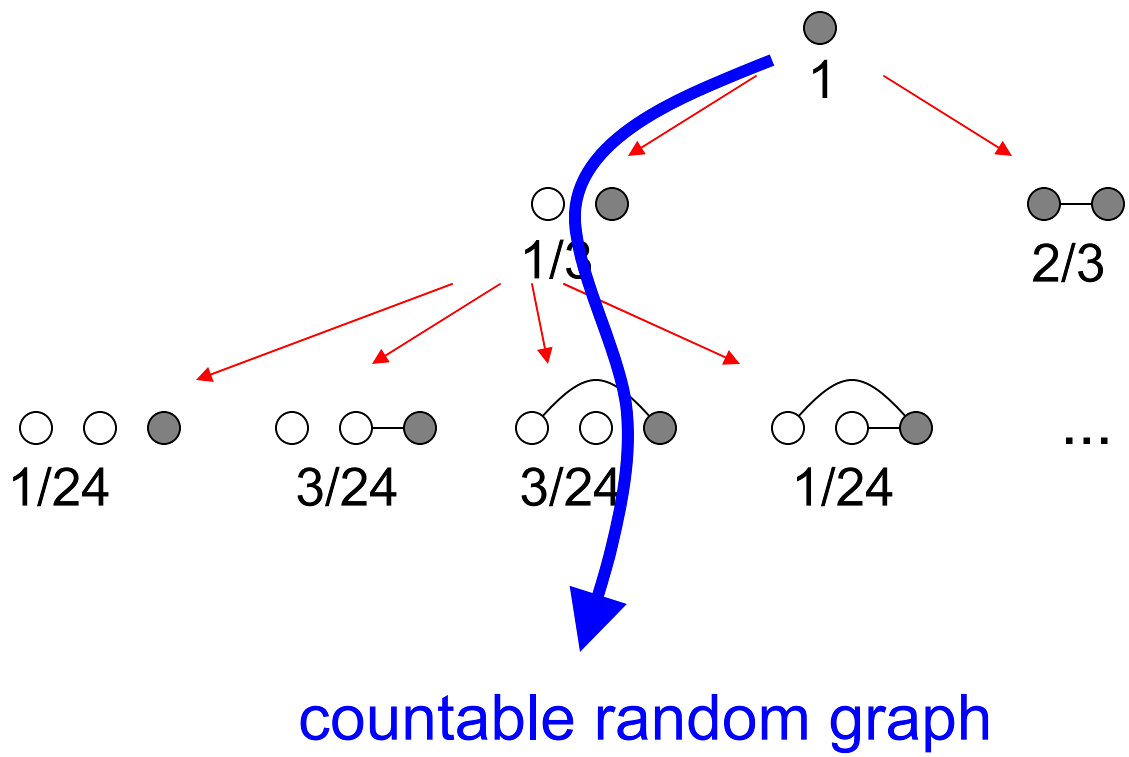
local

(b) for $S = S_1 \dot\cup S_2, G[S_1]$ and $G[S_2]$ are independent.

Every random graph model with (a) and (b) is the limit of models $G[S]$.

Limit objects

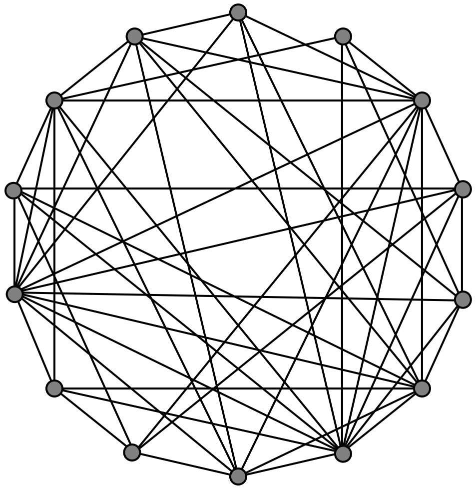
- a consistent local finite random graph model
- a consistent local countable random graph



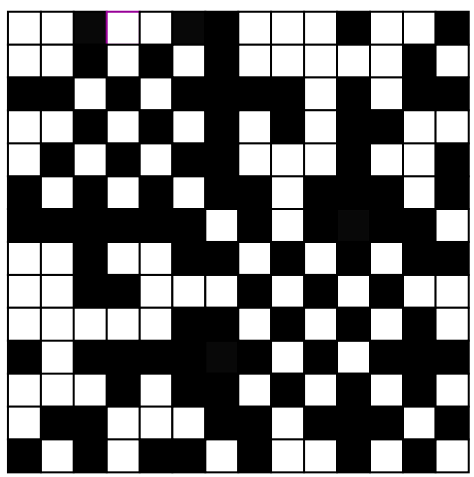
Limit objects

- a consistent local finite random graph model
- a consistent local countable random graph
- a measurable, symmetric function W: $[0,1]^2 \to [0,1]$

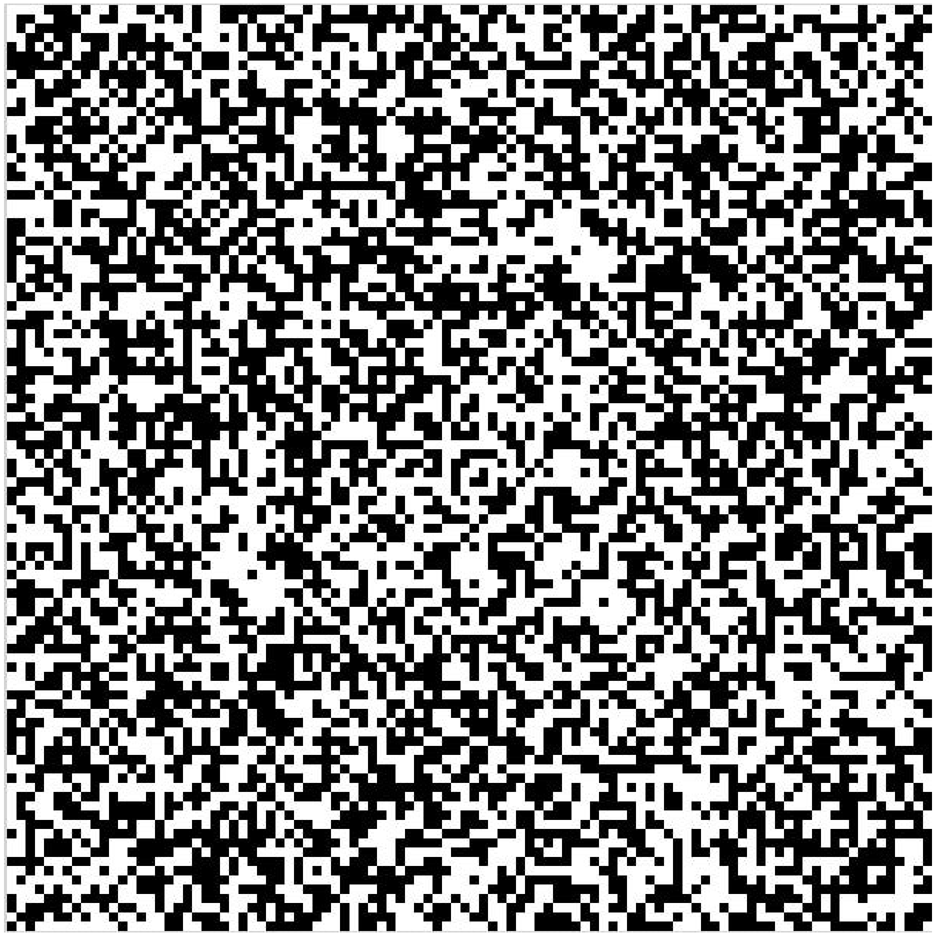
Limit objects

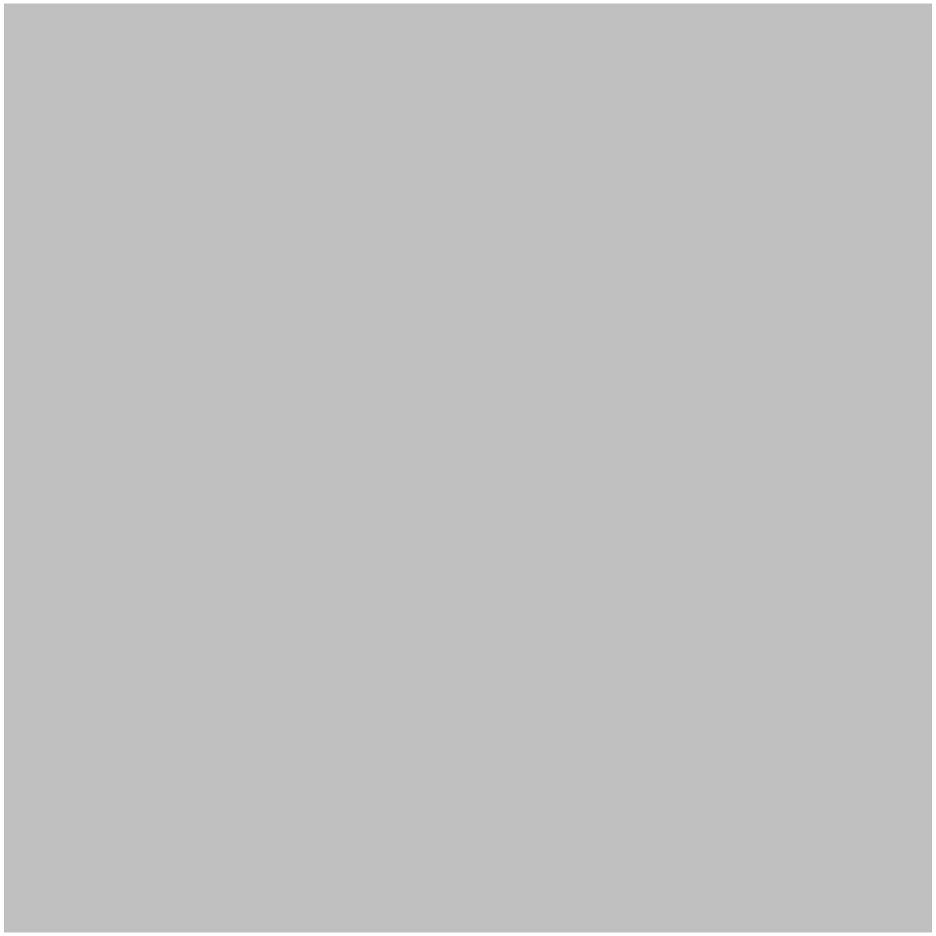


0	0	1	0	0	1	1	0	0	0	1	0	0	1
0	0	1	0	1	0	1	0	0	0	0	0	1	0
1	1	0	1	0	1	1	1	1	0	1	0	1	1
0	0	1	0	1	0	1	0	1	0	1	1	0	0
0	1	0	1	0	1	1	0	0	0	1	0	0	1
1	0	1	0	1	0	1	1	0	1	1	1	0	1
1	1	1	1	1	1	0	1	0	1	1	1	1	0
0	0	1	0	0	1	1	0	1	0	1	0	1	1
0	0	1	1	0	0	0	1	1	1	0	1	0	0
0	0	0	0	0	1	1	0	1	0	1	0	1	0
1	0	1	1	1	1	1	1	0	1	0	1	1	1
0	0	0	1	0	1	1	0	1	0	1	0	1	0
0	1	1	0	0	0	1	1	0	1	1	1	0	1
1	0	1	0	1	1	0	1	0	0	1	0	1	0



Limit objects



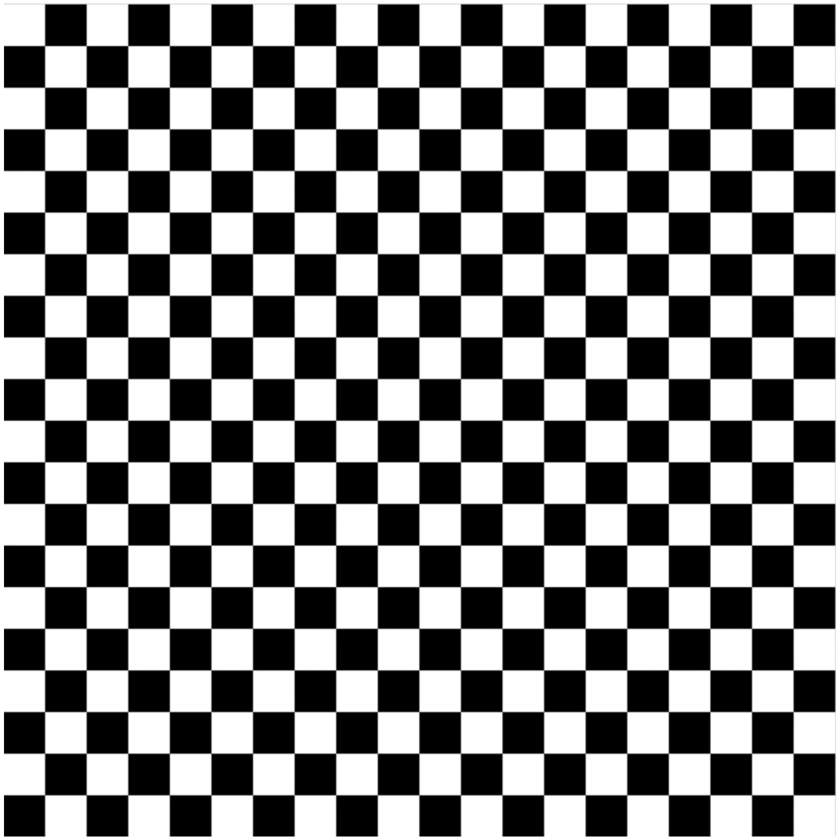


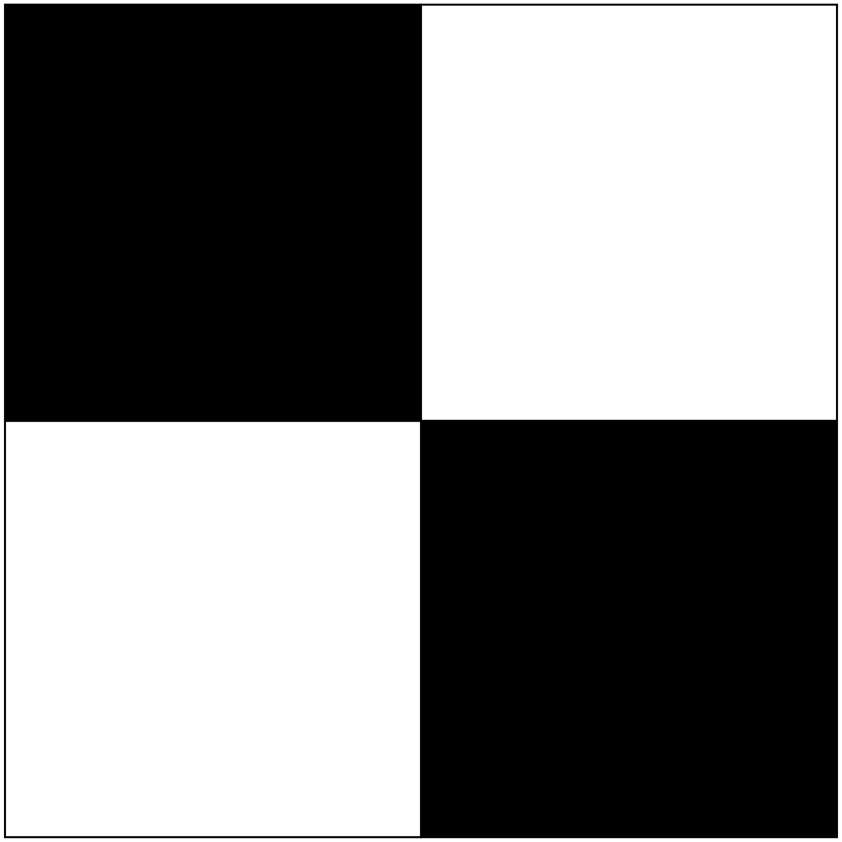
A random graph

with 100 nodes and with 2500 edges

1/2

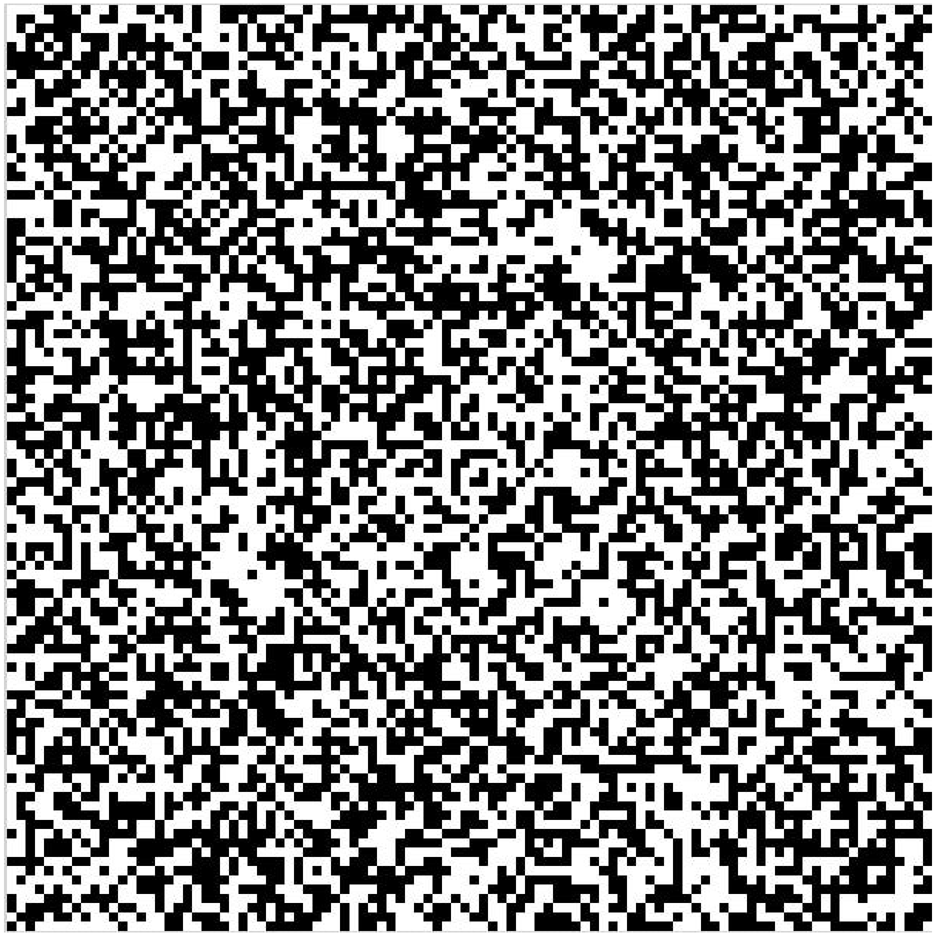
Limit objects

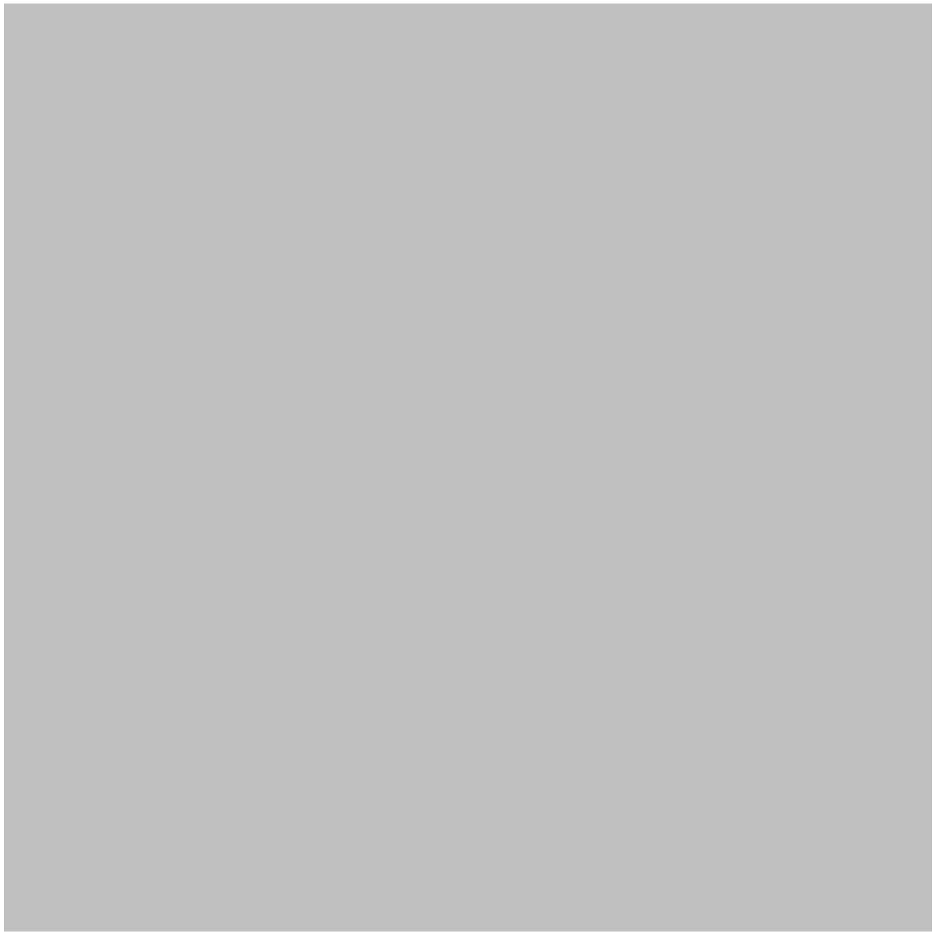




Rearranging the rows and columns

Limit objects





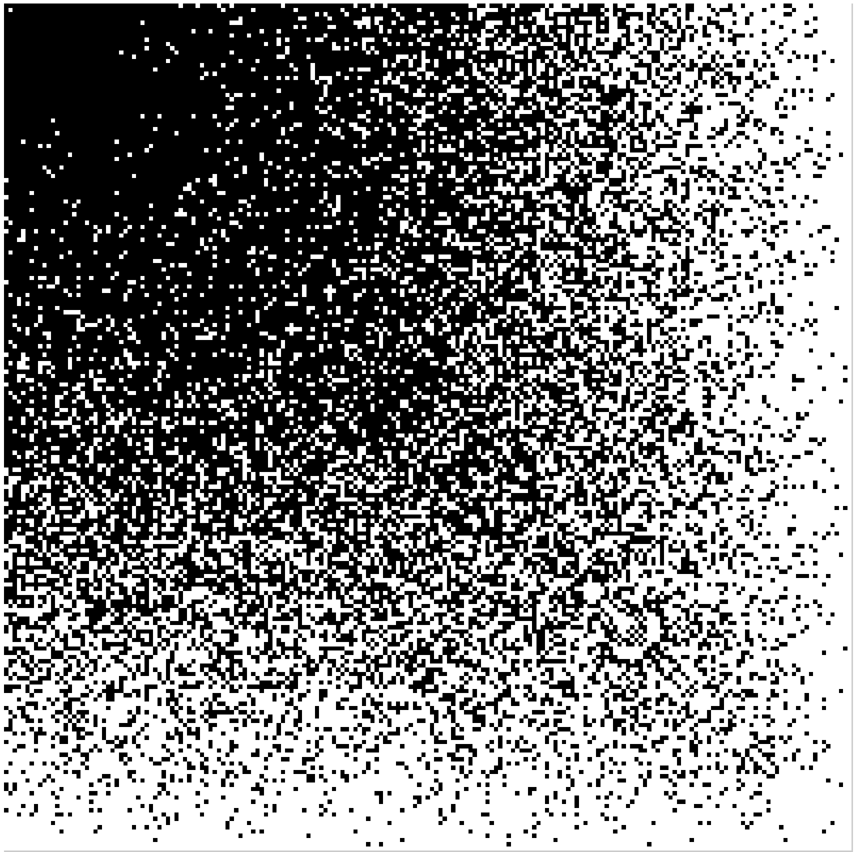
A random graph

with 100 nodes and with 2500 edges

1/2

(no matter how you reorder the nodes)

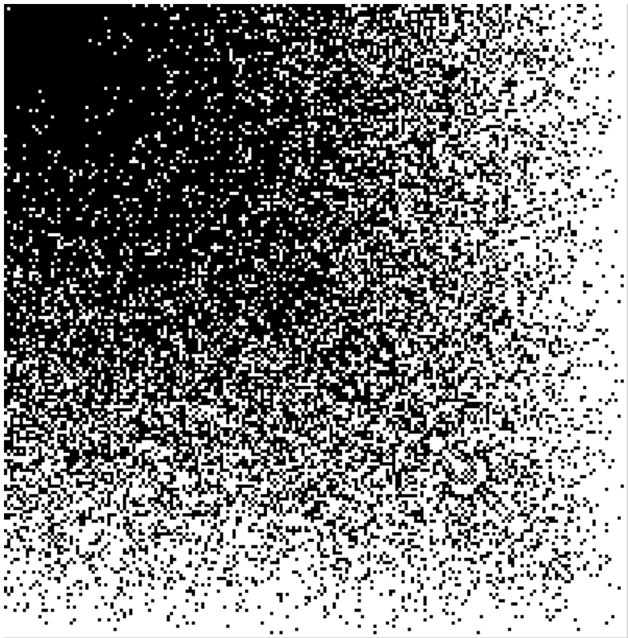
Limit objects



A randomly grown uniform attachment graph with 200 nodes

$$1 - \max(x, y)$$

Limit objects



$$W(x,y) := 1 - \max(x,y)$$

$$t(K_3, G_n) \to \iiint W(x,y)W(y,z)W(x,z)\,dx\,dy\,dz$$

Limit objects

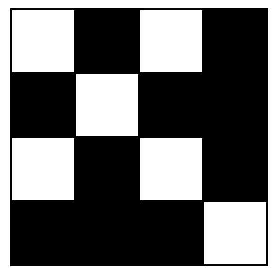
$$\mathcal{W}_0 = \{W : [0,1]^2 \to [0,1] \text{ symmetric, measurable}\}$$

$$t(F,W) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j)\, dx$$

Adjacency matrix of graph G:

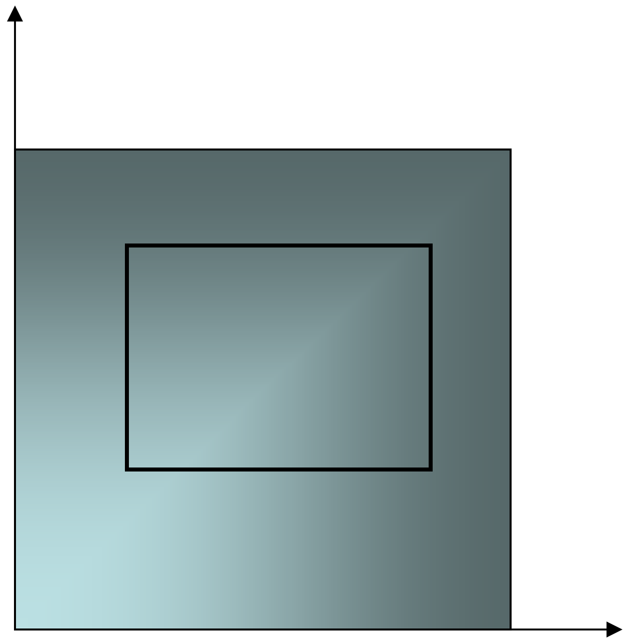
$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Associated function W_G:



$$t(F,G) = t(F,W_G)$$

Distance of functions



$$\delta_\square(W,W') = \inf \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} (W - W') \right|$$

$$\delta_\square(G,G') = \delta_\square(W_G, W_{G'})$$

$(\mathcal{W}_0, \delta_\square)$ is compact.

Equivalent to the Regularity Lemma

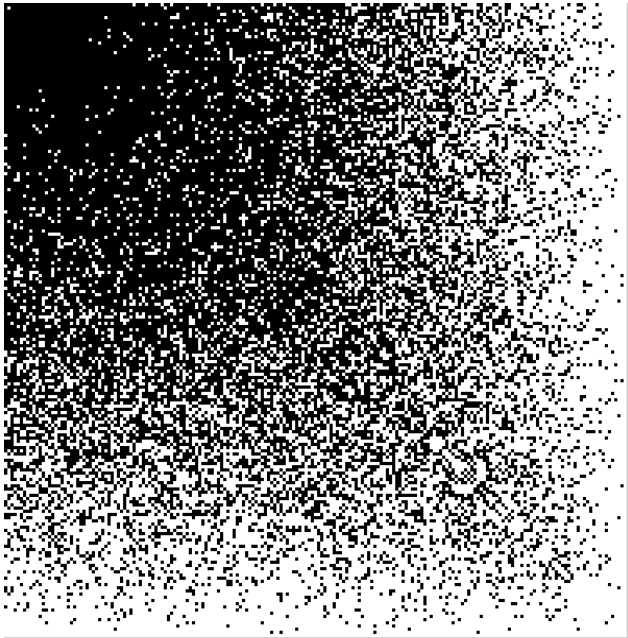
Converging to a function

$G_n \to W$: (i) $\delta(W_{G_n}, W) \to 0$

(ii) $(\forall F)\ \ t(F, G_n) \to t(F, W)$

(i) and (ii) are equivalent.

Limit objects



G_n

$W(x,y)$

$$\delta_\square(G_n, W) \to 0$$

$$\forall F \;\; t(F, G_n) \to t(F, W)$$

Limit objects

For every convergent graph sequence (G_n)

there is a $W \in \mathcal{W}_0$ such that $G_n \to W$.

Conversely, $\forall W \ \exists (G_n)$ such that $G_n \to W$

LL – B. Szegedy

W is essentially unique (up to measure-preserving transform).

Borgs – Chayes - LL

Limit objects

- a consistent local finite random graph model
- a consistent local countable random graph
- a measurable, symmetric function W: $[0,1]^2 \to [0,1]$

Fix $W:[0,1]^2 \to [0,1]$. Let $X_1,\dots,X_n \in [0,1]$ ind uniform.

$$V(\mathbb{G}(n,W)) = \{1,\dots,n\}$$

$$\mathsf{P}\big(ij \in E(\mathbb{G}(n,W))\big) = W(X_i, X_j)$$

← W-random graphs

$$W \equiv 1/2 \quad \Rightarrow \quad \mathbf{G}(n,1/2)$$

Limit objects

- a consistent local finite random graph model
- a consistent local countable random graph
- a measurable, symmetric function *W*: $[0,1]^2 \to [0,1]$
- a multiplicative graph parameter with nonnegative Möbius transform

$$f^\dagger(F) = \sum_{F' \supseteq F} (-1)^{|E(F') \setminus E(F)|} f(F')$$

$$f(F) = \mathsf{P}(F \subseteq G[\mathbf{S}_k]) \qquad f^\dagger(F) = \mathsf{P}(F = G[\mathbf{S}_k])$$

$$t(F,W) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j)\, dx$$

Limit objects

- a consistent local finite random graph model;
- a consistent local countable random graph;
- a measurable, symmetric function W: $[0,1]^2 \to [0,1]$;
- a multiplicative graph parameter with nonnegative Möbius transform;
- a multiplicative, reflection positive graph parameter; (connection matrices are positive semidefinite)

Many applications in extremal graph theory

Limit objects

- a consistent local finite random graph model;
- a consistent local countable random graph;
- a measurable, symmetric function W: $[0,1]^2 \to [0,1]$;
- a multiplicative graph parameter with nonnegative Möbius transform;
- a multiplicative, reflection positive graph parameter;
- A point in the completion of the set of finite graphs with the cut-metric.

Parameter estimation

Graph parameter f is estimable:

$$\forall \varepsilon > 0 \quad \exists k \geq 1 \quad \mathsf{P}(| f(G[\mathbf{S}_k]) - f(G) | > \varepsilon) < \varepsilon.$$

f is estimable

$\Leftrightarrow$

$f(G_n)$ is convergent if (G_n) is convergent

Parameter estimation

f is estimable

$\Leftrightarrow$

(1) $\forall\varepsilon\exists\delta\ \ V(G)=V(G'),\ \ d_\square(G,G')<\delta$

$$\Rightarrow \left|f(G)-f(G')\right|<\varepsilon$$

(2) if $G(m)$ is obtained from G by replacing each node by m copies, then $f(G(m))$ is convergent.

(3) $\forall\varepsilon\exists k\quad |V(G)|>k \quad\Rightarrow\quad |f(G\setminus v)-f(G)|<\varepsilon$

Borgs, Chayes, LL, Sós, Vesztergombi