Random Graph Embeddings

Jason Cantarella[1]
joint with Tetsuo Deguchi[2], Clayton Shonkwiler[3], Erica Uehara[2]

[1]University of Georgia

[2]Ochanomizu University, Physics Department

[3]Colorado State University

Tangled in Knot Theory
ICERM
May 2023

Random Graph Embeddings

Question

Given a graph $\mathfrak{G}$ with $\mathfrak{V}$ vertices and $\mathfrak{E}$ edges, can we construct a random embedding of the graph into $\mathbb{R}^d$ which "respects the graph structure"?

Related to machine learning literature on graph embeddings (`node2vec`, Laplacian eigenmaps), but not exactly the same.

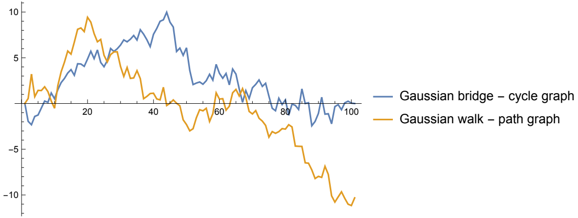
Motivated by physics/chemistry model of Gaussian phantom network (James, Guth, Flory, Eichinger).

Goal is to say as much as possible about a very general class of models and specialize as late as possible.

A motivating example

Definition

A Gaussian *walk* has i.i.d. Gaussian steps. A Gaussian *bridge* is a Gaussian walk which returns to the starting value.



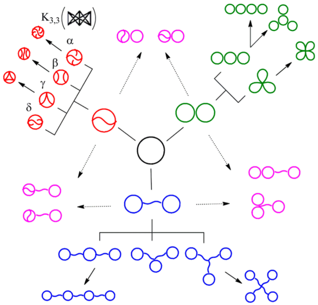
The steps in the Gaussian bridge are not independent.

Goal

Understand dependence structure of edge r.v.s implied by graph type.

Original motivation: “Topological” polymers

Chemists now able to synthesize polymers with various graph types usable quantities:



synthetic topological polymers
(source: Tezuka, 2018)

Graph $\to$ chain complex

Let $\mathfrak{G}$ be a connected, directed graph with $\mathfrak{E}$ edges and $\mathfrak{V}$ vertices.

Definition

The real vector space VC of *vertex chains* is the vector space of (formal) linear combinations of vertices with coefficients $x_i \in \mathbb{R}$:

$$x = x_1 v_1 + \cdots + x_{\mathfrak{V}} v_{\mathfrak{V}}.$$

Definition

The real vector space EC of *edge chains* is the vector space of (formal) linear combinations of edges with coefficients $w_i \in \mathbb{R}$:

$$w = w_1 e_1 + \cdots + w_{\mathfrak{E}} e_{\mathfrak{E}}.$$

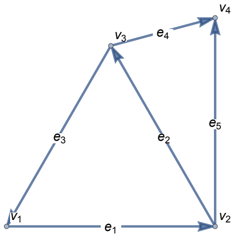
Graph $\to$ chain complex

Definition

The boundary map or incidence matrix $\partial : \mathrm{EC} \to \mathrm{VC}$ is the linear map

$$\partial(e_i) = +1\,\mathrm{head}(e_i) - 1\,\mathrm{tail}(e_i)$$

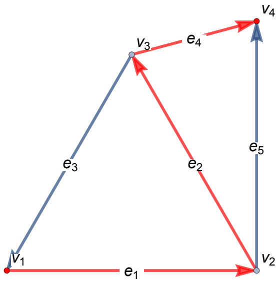
The *chain complex* for $\mathfrak{G}$ is $C_0 = \mathrm{VC} \xrightarrow{\partial} \mathrm{ED} = C_1$.



$$\partial = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Integral chains

We say that a chain is *integral* if the coefficients are integers. Integral edge chains w where $\partial w = v_j - v_i$ have a natural interpretation as paths from v_i to v_j.



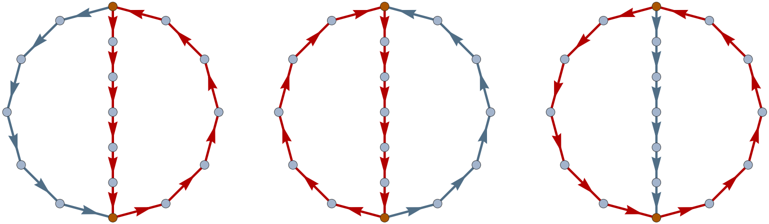
$$
\begin{aligned}
w &= e_1 + e_2 + e_4 \\
\partial w &= (v_2 - v_1) + (v_3 - v_2) + (v_4 - v_3) \\
&= v_4 - v_1
\end{aligned}
$$

Cycles and subspaces

A chain is a set of (real) scalar weights on vertices or edges.

Lemma

A path in $\mathfrak{G}$ *is a cycle* $\iff$ *its integral chain* $w \in \ker \partial \subset \mathrm{EC}$.



Cycle rank $=$ Betti $\#$ $=$ Euler characteristic

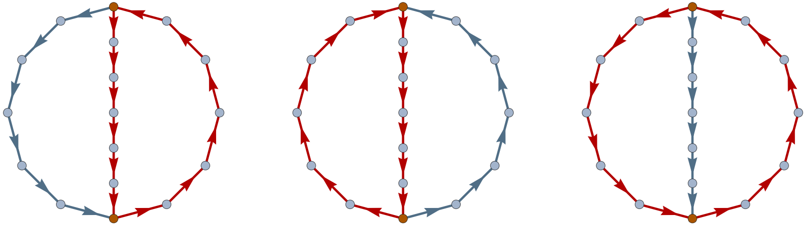
Definition

The subspace $\ker \partial \subset \mathrm{EC}$ is the *first homology group* or *loop space*.

Proposition

$\dim \ker \partial$ *is the cycle rank* $\xi(\mathfrak{G}) = \mathfrak{E} - \mathfrak{V} + 1$ *or first Betti number:*

$$\xi(\mathfrak{G}) = \frac{1}{2} \sum_i (\deg(v_i) - 2) + 1$$



Cycle rank = Betti # = Euler characteristic

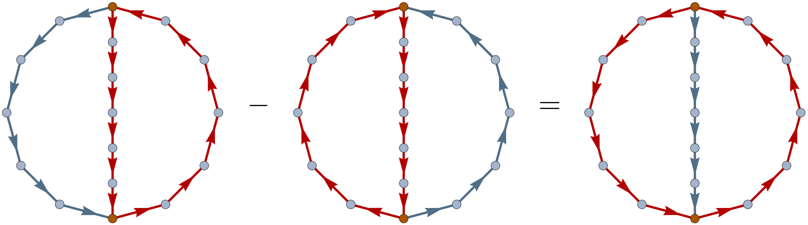
Definition

The subspace $\ker \partial \subset$ EC is the *first homology group* or *loop space*.

Proposition

dim ker ∂ *is the cycle rank* $\xi(\mathfrak{G}) = \mathfrak{E} - \mathfrak{V} + 1$ *or first Betti number:*

$$\xi(\mathfrak{G}) = \frac{1}{2}\sum_i (\deg(v_i) - 2) + 1$$



Cochain spaces

Suggestive observation

Usually build cochain complex by dualizing. We ***can*** *use* $\mathrm{Hom}(-, G)$ *for* ***any*** *abelian group* G*.*

$$
\begin{array}{ccccccccc}
\xleftarrow[\partial]{} & C_{i-1} & \xleftarrow[\partial]{} & C_i & \xleftarrow[\partial]{} & C_{i+1} & \xleftarrow[\partial]{} \\
& \Big\downarrow{\scriptstyle \text{duality}} & & \Big\downarrow{\scriptstyle \text{duality}} & & \Big\downarrow{\scriptstyle \text{duality}} & \\
\xrightarrow{\partial^*} & C^{i-1} & \xrightarrow{\partial^*} & C^i & \xrightarrow{\partial^*} & C^{i+1} & \xrightarrow{\partial^*}
\end{array}
$$

Cochain spaces

Suggestive observation

Usually build cochain complex by dualizing. We ***can*** *use* $\mathrm{Hom}(-, G)$ *for* ***any*** *abelian group* G.

$$\begin{array}{ccccccc}
0 & \xleftarrow[0]{\quad} & C_0 & \xleftarrow[\partial]{\quad} & C_1 & \xleftarrow[0]{\quad} & 0 \\
 & & \Big\downarrow{\scriptstyle \text{duality}} & & \Big\downarrow{\scriptstyle \text{duality}} & & \\
0 & \xrightarrow{0} & C^0 & \xrightarrow{\partial^*} & C^1 & \xrightarrow{0} & 0
\end{array}$$

Cochain spaces

Suggestive observation

Usually build cochain complex by dualizing. We ***can*** *use* $\mathrm{Hom}(-, G)$ *for* ***any*** *abelian group G.*

$$
\begin{array}{ccccccc}
0 & \xleftarrow[0]{} & C_0 & \xleftarrow[\partial]{} & C_1 & \xleftarrow[0]{} & 0 \\
 & & \Big\downarrow{\scriptstyle \mathrm{Hom}(-,\mathbb{R}^d)} & & \Big\downarrow{\scriptstyle \mathrm{Hom}(-,\mathbb{R}^d)} & & \\
0 & \xrightarrow{0} & C^0 & \xrightarrow{\partial^*} & C^1 & \xrightarrow{0} & 0
\end{array}
$$

The cochain spaces

A cochain is a set of vector weights on vertices or edges.

Definition

The real vector space of *vertex cochains* $\mathrm{VP} = \mathrm{Hom}(\mathrm{VC}, \mathbb{R}^n)$ is the space of linear maps $X : \mathrm{VC} \to \mathbb{R}^d$.

Definition

The real vector space of *edge cochains* $\mathrm{ED} = \mathrm{Hom}(\mathrm{EC}, \mathbb{R}^n)$ is the space of linear maps $W : \mathrm{EC} \to \mathbb{R}^d$.

Definition

The coboundary map $\partial^* : \mathrm{VP} \to \mathrm{ED}$ is given by

$$\partial^*(X)(e_i) = X(\partial(e_i)) = X(\mathrm{head}(e_i)) - X(\mathrm{tail}(e_i))$$

cochains $\to$ graph embeddings

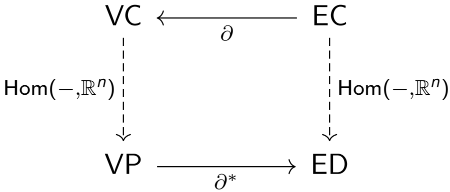
Proposition

Embeddings of $\mathfrak{G}$ in $\mathbb{R}^d$ are a vector space $(X, W) \in \mathrm{VP} \times \mathrm{ED}$ where

$$W = \partial^* X,$$

$X(v_i)$ is the position of v_i,

$W(e_i) = X(\mathrm{head}(e_i)) - X(\mathrm{tail}(e_i))$ is the displacement along e_i.



Cocycles and the failure-to-close map

Definition

The vector space $H^1 = \mathrm{Hom}(\ker \partial, \mathbb{R}^d)$ is the *first cohomology* or *coloop* space.

Lemma

If $L \in H^1$ and we have a cycle in $\mathfrak{G}$ with integral chain $w \in H_1$, then $L(w)$ is the sum of the displacements over the cycle. We call it the failure to close of the cycle w.

Definition

The dual of the inclusion map $i : H_1 \hookrightarrow C_1$ is given by the restriction map $i^* : C^1 \to H^1$. We call i^* the *failure to close* map ftc.

What it means to condition on graph type

Theorem

We have $W = \partial^* X \iff W \in \operatorname{ker}\operatorname{ftc}$. *If* $(\partial^*)^+$ *is the Moore-Penrose pseudoinverse, it is true that*

$\operatorname{im}(\partial^*)^+$ *is an isomorphism when restricted to* $\operatorname{ker}\operatorname{ftc}$*.*

$\operatorname{im}(\partial^*)^+ = \operatorname{im}(\partial^*)^+(\operatorname{ker}\operatorname{ftc})$.

$X \in \operatorname{im}(\partial^*)^+ \iff X(\sum v_i) = 0$.

We call these centered embeddings of $\mathfrak{G}$*.*

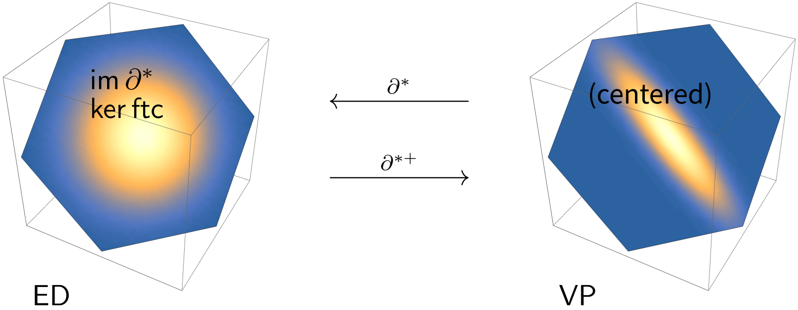
Corollary

A random $W \in \mathrm{ED}$ *is compatible with the graph type* $\iff W \in \operatorname{ker}\operatorname{ftc}$*.*

Example: James-Guth-Flory Theory

Definition

The *phantom network* embedding of $\mathfrak{G}$ is given by choosing W from a standard Gaussian on ED conditioned on $W = \ker \mathrm{ftc}$ and pushing forward this probability distribution to VC by $(\partial^*)^+$.



Early geometric/topological learning feature.

Fact

Expected radius of gyration of graph embedding given by $\frac{d}{\mathfrak{V}} \operatorname{tr} L^{+} = d \operatorname{Kf}(\mathfrak{G})$ *where* $\operatorname{Kf}(\mathfrak{G})$ *is called the Kirchhoff index.*

The Kirchhoff index was defined to generalize the Wiener index for trees, which was a heuristic feature used (by hand!) in 1947 to predict paraffin boiling points to within 0.4° C.

Jan., 1947 STRUCTURAL DETERMINATION OF PARAFFIN BOILING POINTS

[CONTRIBUTION FROM DEPARTMENT OF CHEMISTRY, BROOKLYN COLLEGE]

Structural Determination of Paraffin Boiling Points

BY HARRY WIENER[1]

Suggestive observation (Estrada-Hatano, 2010)

The radius of gyration interpretation explains why the Kirchhoff index is so successful at predicting chemical properties.

Conditioning in elementary probability

Definition

A probability measure $\lambda = f(W)\ \mathrm{dVol}_{\mathrm{ED}}$ on ED with pdf $f(W)$ is *admissible* if the pushfoward measure $\mu = \mathrm{ftc}\,\lambda$ has a density function $g(L)$ on H^1 which is continuous everywhere and nonzero at 0.

We can define a family of conditional probabilities λ_L for W chosen from λ conditioned on the hypothesis that $\mathrm{ftc}(W) = L$ by observing that

$$g(L) = \int_{W \in \mathrm{ftc}^{-1}(L)} f(W)\ \mathrm{dVol}_{\mathrm{ftc}^{-1}(L)}$$

and constructing (for a.e. L and $g(L) \neq 0$)

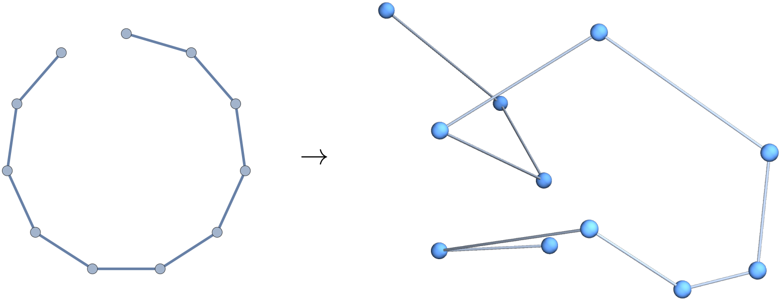
$$\lambda_L = \frac{f(W)}{g(L)}\ \mathrm{dVol}_{\ker \mathrm{ftc}} = \frac{\text{joint density}}{\text{marginal density}}\ \mathrm{dVol}_{\ker \mathrm{ftc}}$$

Sampling Random Graph Embeddings

Compute pseudoinverse matrix ∂^{*+} : ED $\to$ VP.

Sample W from conditional probability λ_0.

Construct vertex positions $X = \partial^{*+} W$.

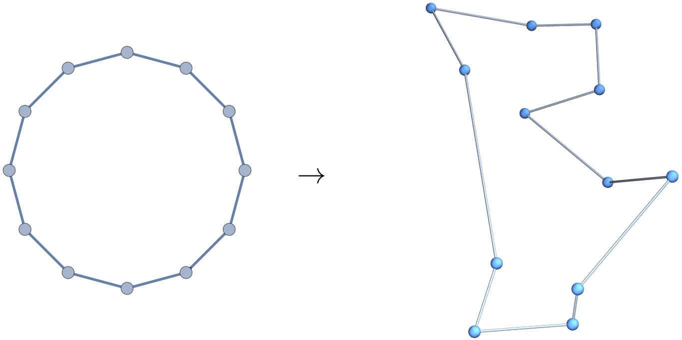


Sampling Random Graph Embeddings

Compute pseudoinverse matrix ∂^{*+} : ED $\to$ VP.

Sample W from conditional probability λ_0.

Construct vertex positions $X = \partial^{*+} W$.

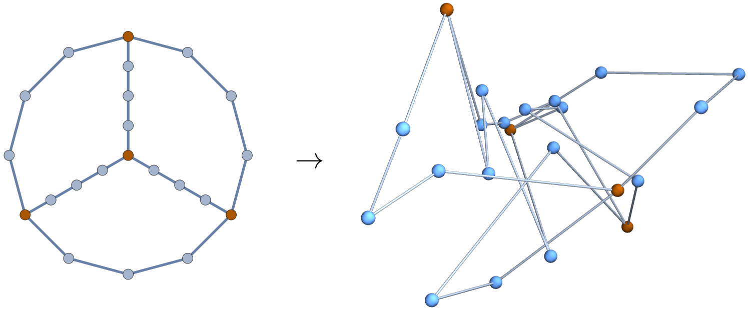


Sampling Random Graph Embeddings

Compute pseudoinverse matrix ∂^{*+} : ED $\to$ VP.

Sample W from conditional probability λ_0.

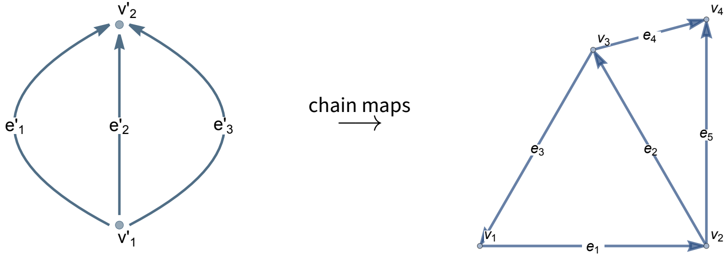
Construct vertex positions $X = \partial^{*+} W$.



Chain maps and simpler graphs

Definition

Given two graphs $\mathfrak{G}$ and $\mathfrak{G}'$, we say that linear maps $f_0 : \mathrm{VC} \to \mathrm{VC}'$ and $f_1 : \mathrm{EC} \to \mathrm{EC}'$ are *chain maps* if $\partial f_1 = f_0 \partial'$.



$$f_0(v_1') = v_2, f_0(v_2') = v_3$$

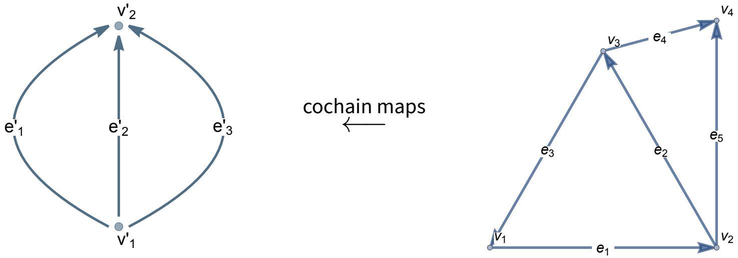
$$f_1(e_1') = -e_1 - e_3, f_1(e_2') = e_2,$$

$$f_1(e_3') = e_5 - e_4$$

Cochain maps and simpler graphs

Definition

Given two graphs $\mathfrak{G}$ and $\mathfrak{G}'$, we say that linear maps $f_0^* : \mathrm{VP}' \to \mathrm{VP}$ and $f_1^* : \mathrm{ED}' \to \mathrm{ED}$ are *cochain maps* if $(\partial')^* f_0^* = f_1^* \partial^*$.



$$(f_0^* X)(v_1') = X(v_2), (f_0^* X)(v_2') = X(v_3)$$
$$(f_1^* W)(e_1') = -W(e_1 + e_5), (f_1^* W)(e_2') = W(e_2),$$
$$(f_1^* W)(e_3') = W(e_5 - e_4)$$

Which chain maps are ok?

Proposition

Suppose that f_0 and f_1 are chain maps, f_1 is injective and $\dim H_1 = \dim H_1'$. *There is a unique isomorphism ϕ^* giving a commutative*

$$\begin{array}{ccccc} \mathrm{VP} & \xrightarrow{\partial^*} & \mathrm{ED} & \xrightarrow{\mathrm{ftc}} & H^1 \\ \downarrow f_0^* & & \downarrow f_1^* & & \downarrow \phi^* \\ \mathrm{VP}' & \xrightarrow{(\partial')^*} & \mathrm{ED}' & \xrightarrow{\mathrm{ftc}'} & (H^1)' \end{array}$$

That is, the chain map takes cocycles to cocycles.

Main theorem: Law of Iterated Expectations

Theorem

Given an admissible probability measure λ on ED*, let $\lambda' = f_1^* \lambda$, $L' = \phi^* L$, and $W' = f_1^* W$. We may construct conditional probabilities*

λ_L *for λ, based on the map* ftc*,*

$\lambda_{W'}$ *for λ, based on the map f_1^*,*

$\lambda'_{L'}$ *for λ', based on the map* ftc$'$*.*

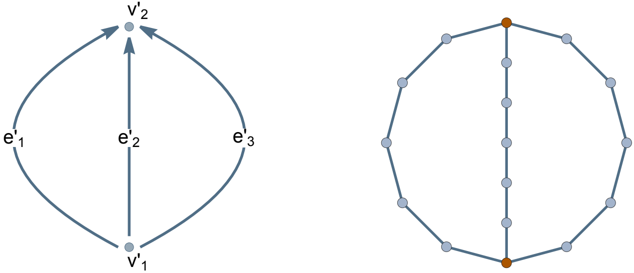
and write $\lambda_L = (\lambda'_{L'})^{W'} \lambda_{W'}$.

"For any r.v. f, the c.e. of $f(W)$ given $\mathrm{ftc}(W) = L$ is equal to c.e. of (the c.e. of f given $f_1^*(W)$) given $\mathrm{ftc}'(W') = L'$."

Application: Subdivision graphs

Definition

We say $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ if there is a partition of $\{1, \ldots, \mathfrak{E}\}$ into $\mathfrak{E}'$ subsets $\mathcal{S}_i$ and chain maps f_0, f_1 with $f_1(e'_i) = \sum_{j \in \mathcal{S}_i} e_j$.



Conditional independence in subdivisions

Proposition

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$, and λ is a product of independent distributions λ^j on the spaces $\mathsf{ED}^j = \mathsf{Hom}(e_j, \mathbb{R}^d)$, then λ is the product of independent distributions

$$\lambda^{\mathcal{S}_i} = \prod_{j \in \mathcal{S}_i} \lambda^j$$

Further, each conditional probability

$$\lambda_{W'} = \prod_i \lambda^{\mathcal{S}_i}_{W'_i}$$

where $\lambda^{\mathcal{S}_i}_{W'_i}$ is the probability of the edge displacements in $\mathcal{S}_i$ conditioned on the hypothesis that their sum is W'_i.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

Compute $\lambda' = f_1^* \lambda$, including pdf, variance.

Corollary

Subdivisions with large enough groups are approximately Gaussian.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

Define $\hat{\lambda}'$ to be the Gaussian with matching mean, variance.

Corollary

Subdivisions with large enough groups are approximately Gaussian.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

Choose W' from the conditional probability $\hat{\lambda}'_0$.

Corollary

Subdivisions with large enough groups are approximately Gaussian.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

Reweight W' by $\lambda'(W')/\hat{\lambda}'(W')$.

Corollary

Subdivisions with large enough groups are approximately Gaussian.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

For each i, sample W_j for $j \in \mathcal{S}_i$ according to $\lambda^{\mathcal{S}_i}_{W'_i}$.

Corollary

Subdivisions with large enough groups are approximately Gaussian.

Sampling subdivision graphs

If $\mathfrak{G}$ is a subdivision of $\mathfrak{G}'$ and λ is an admissible measure on ED with independent edges:

Assemble W_j into final sample.

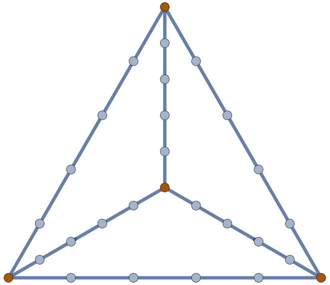
Corollary

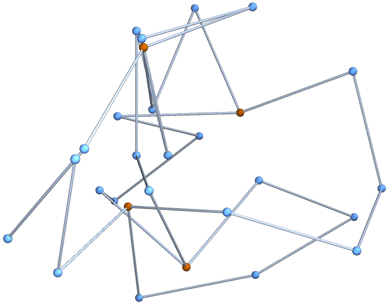
Subdivisions with large enough groups are approximately Gaussian.

Freely-jointed networks

Definition

If the measure μ on ED is the submanifold measure on the product of unit spheres $(S^2)^{\mathfrak{E}} \subset \mathrm{ED} = (\mathbb{R}^3)^{\mathfrak{E}}$, we call the resulting model a *freely jointed network*.

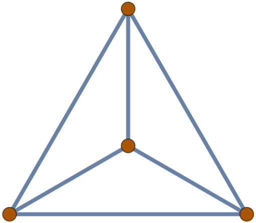


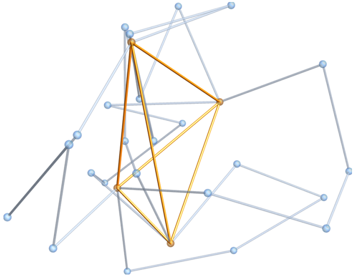


Freely-jointed networks

Definition

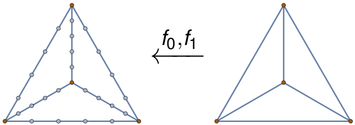
If the measure μ on ED is the submanifold measure on the product of unit spheres $(S^2)^{\mathfrak{E}} \subset \mathrm{ED} = (\mathbb{R}^3)^{\mathfrak{E}}$, we call the resulting model a *freely jointed network*.



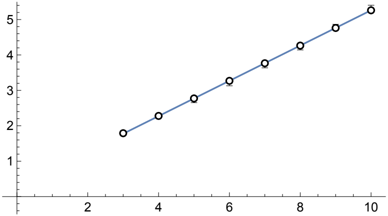


Junction-junction distance

With the obvious chain maps:



can compute μ' on ED$'$ explicitly. Junction-junction distances are explicit 6-d numerical integrals inside ED$'$.



Comparison with Markov-chain experiments

What happens as subdivisions $\to \infty$?

Definition

The *normalized* graph Laplacian $\mathcal{L}(\mathfrak{G})$ is given by

$$\mathcal{L}_{ij} = \begin{cases} 1 - \frac{2 \times \#\text{ loop edges}}{\text{degree}(v_i)}, & \text{if } i = j, \\ -\frac{k}{\sqrt{\text{degree}(v_i)\,\text{degree}(v_j)}}, & \text{if } v_i, v_j \text{ joined by } k \text{ edges}, \\ 0, & \text{otherwise.} \end{cases}$$

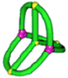
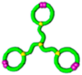
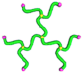
Theorem (with Deguchi, Shonkwiler, Uehara)

$$\lim_{n \to \infty} \frac{1}{\mathfrak{V}(\mathfrak{G}_n)} \mathcal{E}(R_g^2(\mathfrak{G}_n)) = \frac{1}{\mathfrak{E}(\mathfrak{G})^2} \left(\operatorname{Tr} \mathcal{L}^+(\mathfrak{G}) + \frac{1}{3} \operatorname{Loops}(\mathfrak{G}) - \frac{1}{6} \right)$$

Proof.

Superposition of solutions to Poisson problems. $\square$

Tezuka polymer predictions

$\mathfrak{G}$	$\mathcal{E}(R_g^2(\mathfrak{G}_n))$ (with $v = \mathfrak{V}(\mathfrak{G}_n)$)	$\lim_{n\to\infty} \frac{1}{\mathfrak{V}(\mathfrak{G}_n)}\mathcal{E}(R_g^2(\mathfrak{G}_n))$
	$\frac{17v^3+60v^2-261v+108}{486v^2}$	$\frac{17}{486}$
	$\frac{107v^3+270v^2-933v+340}{2430v^2}$	$\frac{107}{2430}$
	$\frac{109v^3+372v^2-1305v+540}{2430v^2}$	$\frac{109}{2430}$
	$\frac{31v^3+78v^2-177v+68}{486v^2}$	$\frac{31}{486}$
	$\frac{43v^3+108v^2-165v+68}{486v^2}$	$\frac{43}{486}$
	$\frac{49v^3+96v^2-177v+32}{486v^2}$	$\frac{49}{486}$

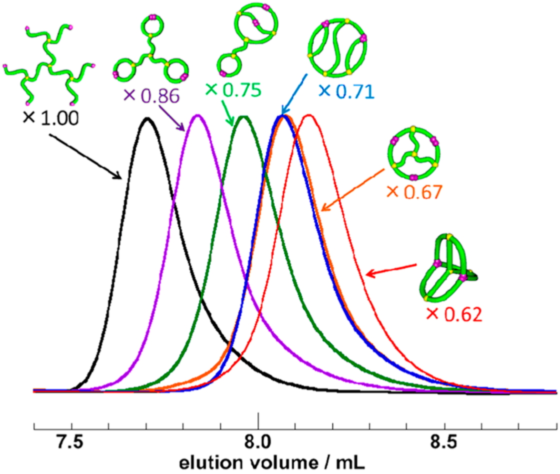
Experimental measurement of relative size

Size-exclusion Chromatography (SEC) apparatus

Tezuka lab

(Source: Cantarella, 2018)

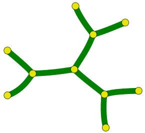
Experimental measurement of relative size



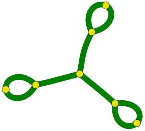
(Suzuki, Yamamoto, Tezuka, 2014)

Comparison of theory and experiment

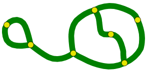
The relative $\lim_{n\to\infty} \frac{1}{\mathfrak{V}(\mathfrak{G}_n)} \mathcal{E}(R_g^2(\mathfrak{G}_n))$ values are:



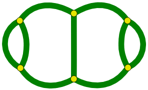
$(1, 1.00)$



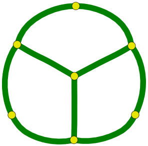
$(\frac{43}{49}, 0.86)$



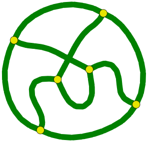
$(\frac{31}{49}, 0.75)$



$(\frac{109}{245}, 0.71)$

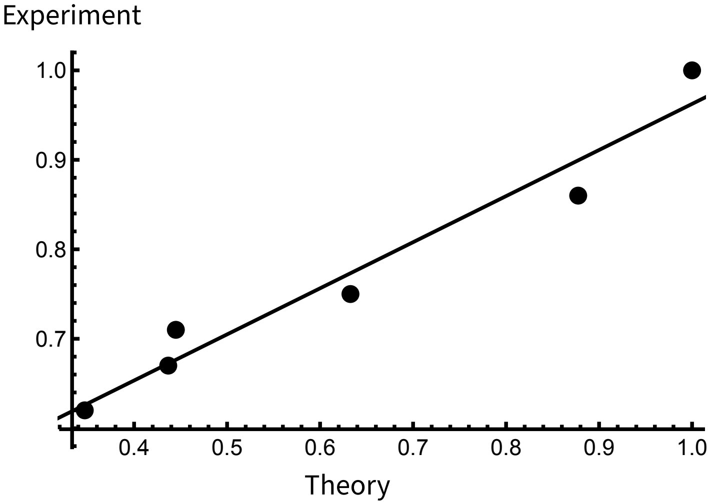


$(\frac{107}{245}, 0.67)$

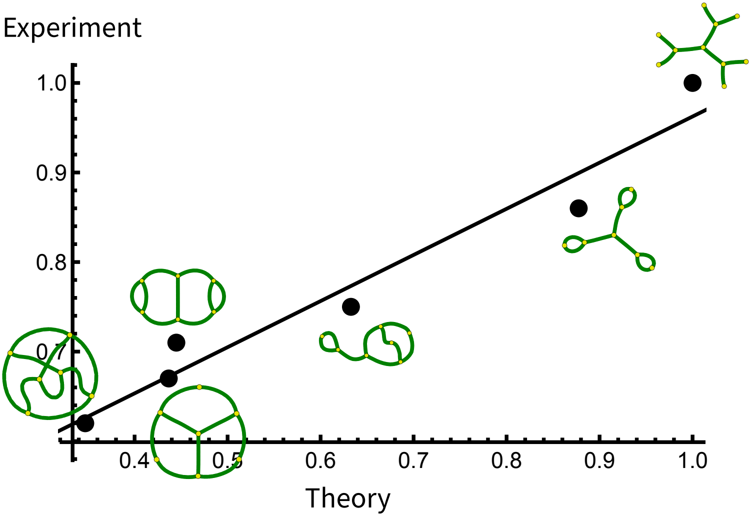


$(\frac{17}{49}, 0.62)$

Comparison of theory and experiment: II



Comparison of theory and experiment: II

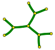
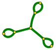
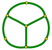


Comparison of theory and simulation

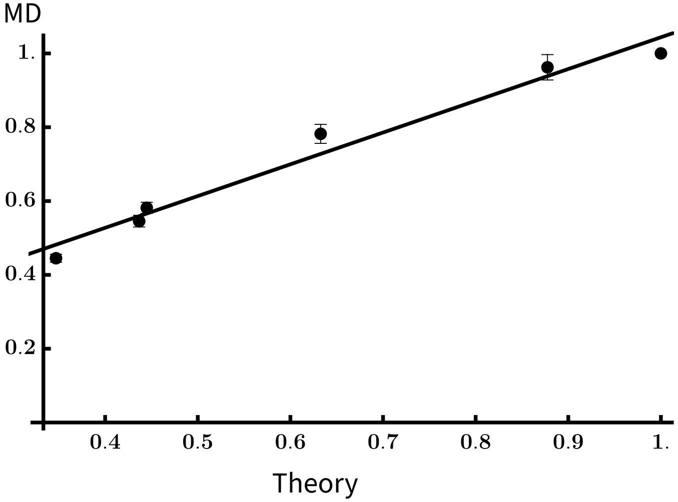
We performed molecular dynamics simulations using LAMMPS on the TSUBAME supercomputer at Tokyo Tech. These included self-avoidance so the radii of gyration fit to

$$\mathcal{E}(R_g^2; \mathfrak{G}_n) = C_{\mathfrak{G}} \mathfrak{V}(\mathfrak{G}_n)^{1.176} + \Delta_{\mathfrak{G}}$$

and we could estimate $g(\mathfrak{G}_\infty, \mathfrak{G}_\infty^{\text{tree}})^{\text{MD}} = C_{\mathfrak{G}} / C_{\text{tree}}$.

$\mathfrak{G}$	MD	theory	$\mathfrak{G}$	MD	theory
	1.0	1		0.962 ± 0.034	$^{43}/_{49}$
	0.782 ± 0.026	$^{31}/_{49}$		0.582 ± 0.015	$^{109}/_{245}$
	0.546 ± 0.016	$^{107}/_{245}$		0.445 ± 0.011	$^{17}/_{49}$

Comparison of theory and simulation II



Thank you for inviting me!

Radius of Gyration, Contraction Factors, and Subdivisions of Topological Polymers, Cantarella, Shonkwiler, Deguchi, Uehara, arXiv:2004.06199

Random graph embeddings with general edge potentials Cantarella, Shonkwiler, Deguchi, Uehara, arXiv:2205.09049

Sampling freely jointed networks Cantarella, Shonkwiler, Schumacher, In preparation.

Also thanks to our funding agencies:

Simons Foundation/SFARI 524120, J.C.
Simons Foundation/SFARI 354225, C.S.
Japan Science and Technology Agency (CREST Grant Number JPMJCR19T4, Deguchi Lab) and JSPS KAKENHI Grant Number JP17H06463.