Homology of Colored Graphs

A Framework for Type-Aware Network Topology

Andrew Belz & Pearson Morris

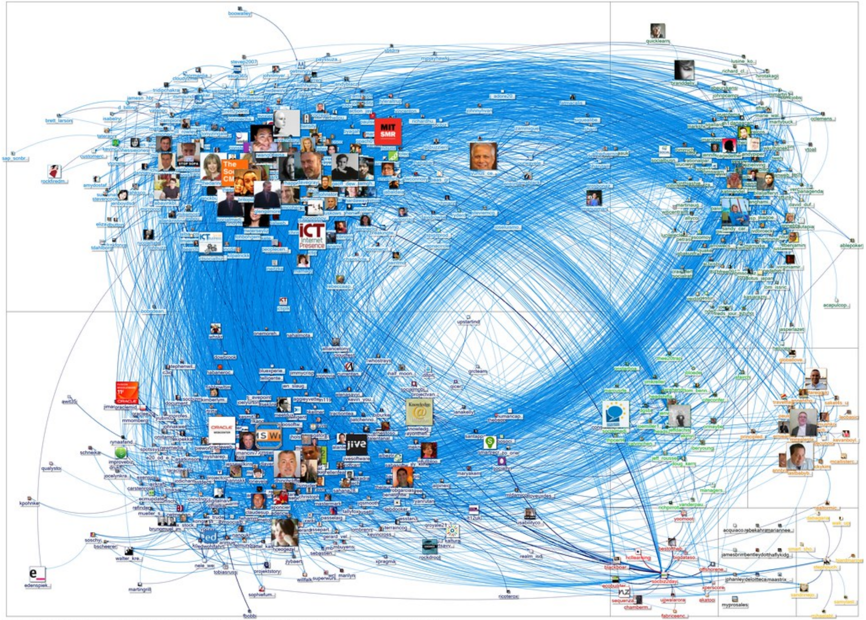
Department of Mathematics
Brigham Young University

Why We Care About Graphs

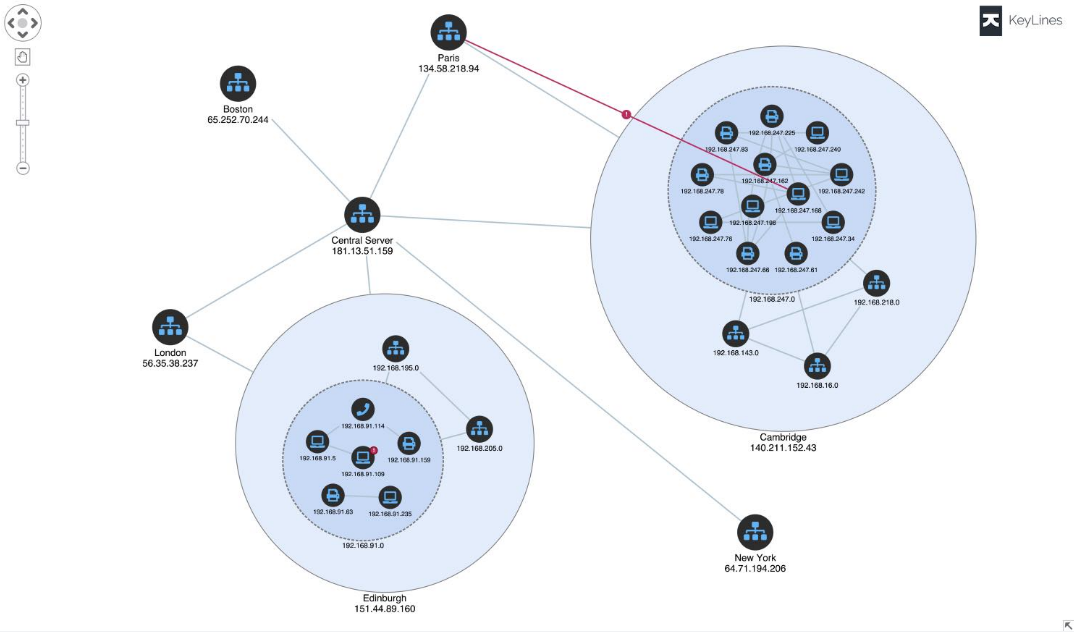
Many systems can be modeled as *entities* and their *relationships*.

- Graphs give us a common language for:
  - social networks
  - communication/cyber networks
  - trade and transportation systems
  - biological and neural networks
- Compare different systems with the same tools.

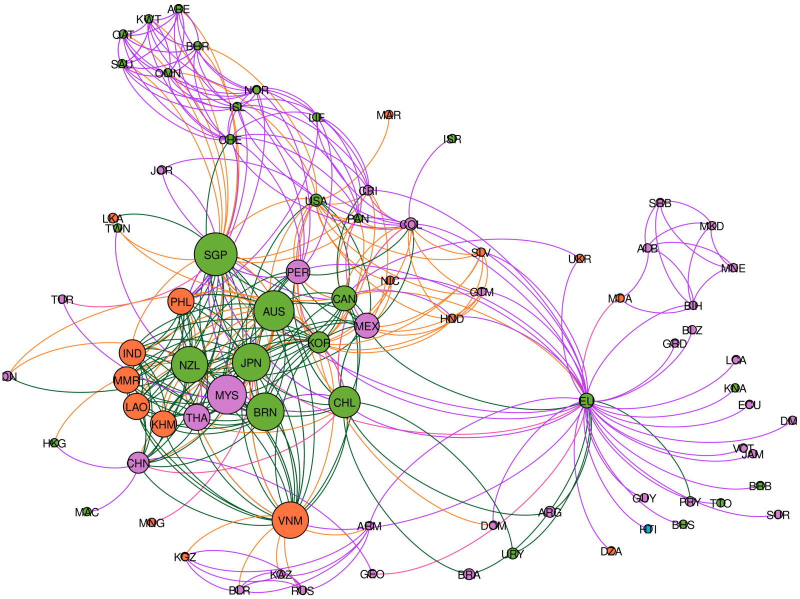
Social media network connections among Twitter users



Created with NodeXL (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)



KeyLines
Paris
134.58.218.94
Boston
65.252.70.244
Central Server
181.13.51.159
London
56.35.38.237
192.168.247.225
192.168.247.83
192.168.247.240
192.168.247.162
192.168.247.78
192.168.247.242
192.168.247.168
192.168.247.198
192.168.247.76
192.168.247.34
192.168.247.66
192.168.247.61
192.168.247.0
192.168.218.0
192.168.143.0
192.168.16.0
Cambridge
140.211.152.43
192.168.195.0
192.168.91.114
192.168.91.5
192.168.91.109
192.168.91.159
192.168.205.0
192.168.91.63
192.168.91.235
192.168.91.0
Edinburgh
151.44.89.160
New York
64.71.194.206


ARE
KWT
QAT
BHR
SAU
OMN
NOR
ISL
LIE
MAR
ISR
CHE
JOR
CRI
USA
PAN
COL
LKA
TWN
SGP
PER
SLV
UKR
SRB
MKD
ALB
MNE
NIC
GTM
MDA
BIH
TUR
PHL
AUS
CAN
HND
MEX
BLZ
GRD
KOR
IND
NZL
JPN
LCA
MMR
KNA
MYS
ECU
CHL
EU
LAO
THA
KHM
BRN
VCT
JAM
HKG
CHN
BRB
GUY
PRY
TTO
ARG
SUR
MAC
MNG
VNM
ARM
DOM
HTI
BHS
URY
DZA
KGZ
KAZ
GEO
BRA
BLR
RUS

What a Graph Captures Well

Graphs

A graph $G = (V, E)$ records which pairs of vertices are connected.

- Classic graph features describe structure:
  - degree
  - connected components
  - shortest paths
  - centrality

Why Pairwise Structure Is Not Enough

- A graph only records **pairwise** interactions.
- Graphs can have several node **types**
- Sometimes we care about **higher-order structure**

Motivation

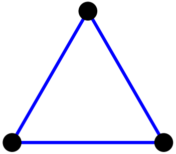
Statistically analyze the higher-order structure of a graph with many different types.

Cliques

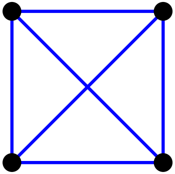
Definition

A *clique* is a maximally connected subgraph.

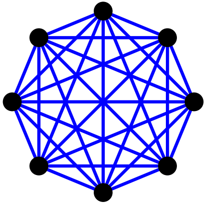
2-clique



3-clique

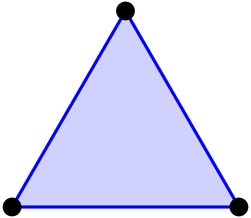


4-clique

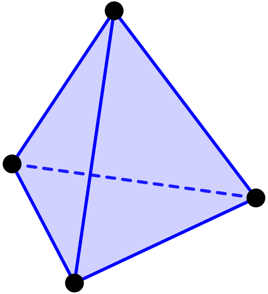


8-clique

The Simplex: an n-Dimensional Triangle

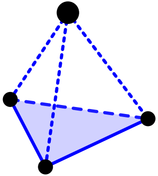


2-simplex
3-clique
(Triangle)

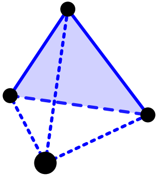


3-simplex
4-clique
(Tetrahedron)

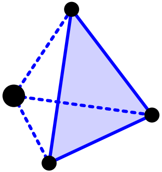
The Boundary of a Simplex Δ^3



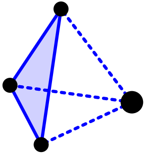
Bottom Face: Δ_1^2



Back Face: Δ_2^2

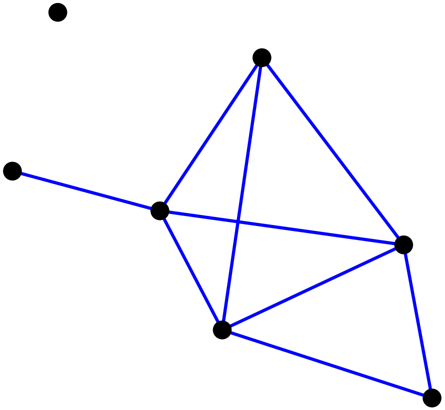


Right Face: Δ_3^2

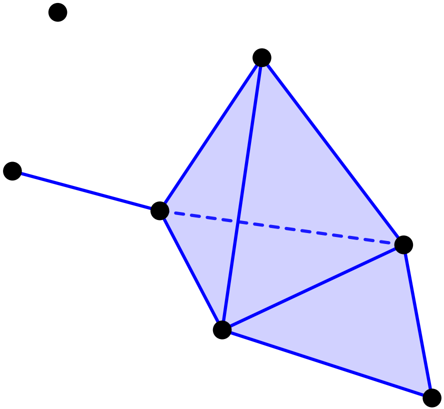


Left Face: Δ_4^2

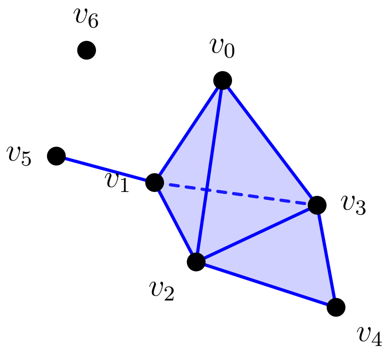
Building a Simplicial Complex from a Graph



Building a Simplicial Complex from a Graph

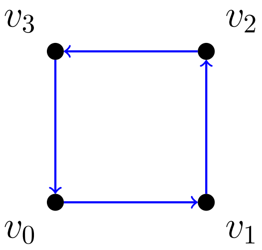


The Boundary Map of a Simplicial Complex



$$\begin{aligned}\partial(\Delta_1^3) &= (1)\Delta_1^2 + (1)\Delta_2^2 + (1)\Delta_3^2 + (1)\Delta_4^2 + (0)\Delta_5^2 \\ &= \begin{bmatrix}\Delta_1^2 & \Delta_2^2 & \Delta_3^2 & \Delta_4^2 & \Delta_5^2\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1 & 0\end{bmatrix}^T\end{aligned}$$

Cycles



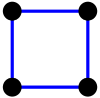
$$C = [v_0, v_1] + [v_1, v_2] + [v_2, v_3] + [v_3, v_0].$$

$$\partial(C) = (v_1 - v_0) + (v_2 - v_1) + (v_3 - v_2) + (v_0 - v_3) = 0.$$

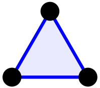
Cycles

Definition

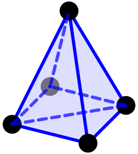
An n-*cycle* is an n-chain that is equal to zero.



Empty 1-Cycle
A "loop" of edges that is *not* filled in.



Filled 1-Cycle
A "loop" that is the boundary of a 2-simplex.



Empty 2-Cycle
A shell of 2-simplices enclosing an empty void.

Homology Groups

Definition

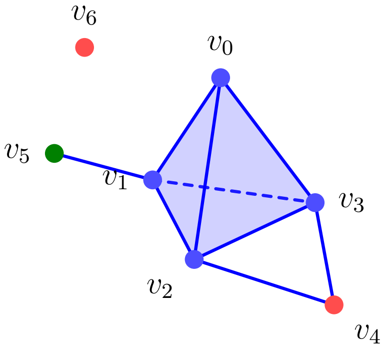
The n^{th} *homology group*, H_n, is the set of all n-cycles that are not the boundaries of an $(n+1)$-simplex.

$$H_n = \frac{\ker(\partial_n)}{\operatorname{im}(\partial_{n+1})}.$$

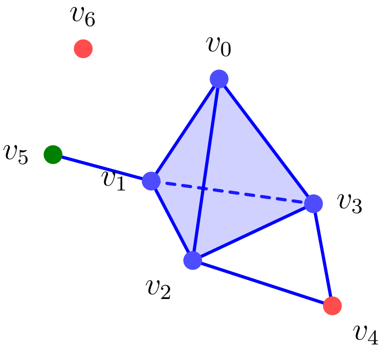
Switching to Multi-Colored Graphs

$$G \mapsto \text{Cliques} \mapsto C_k \mapsto \partial_k \text{ matrices} \mapsto \beta_k$$

Multi-Colored Graphs (Type-Aware)



Enumerating Cliques

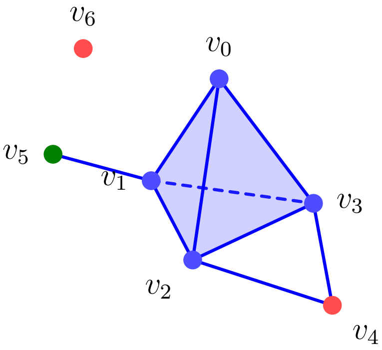


$$\mathcal{B}_0 = \{[v_0], [v_1], [v_2], [v_3], [v_4], [v_5], [v_6]\}$$

$$= \left\{ \begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix} \right\}$$

$$C_0 = span_{\mathbb{F}_2}(\mathcal{B}_0)$$

Boundary Matrices

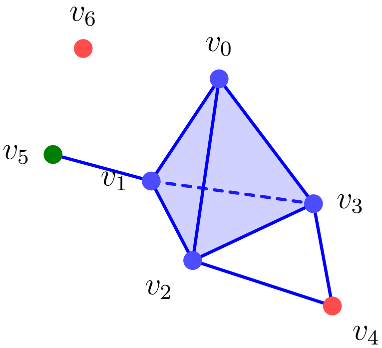


Where: $m = dim(C_0), \ n = dim(C_1)$

$$\partial_1 = D_{m \times n}$$

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

Boundary Matrices



Where: $p = \dim(C_1), \quad q = \dim(C_2)$

$$\partial_2 = D_{p\times q}$$

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

Betti Numbers

Definition

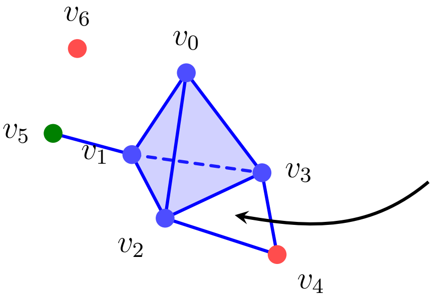
The n^{th} *Betti number*, β_n, is the dimension of H_n (i.e. linearly independent "holes").

$$\dim(H_n) = \text{nullity}(\partial_n) - \text{rank}(\partial_{n+1}).$$

Betti Number Calculations

Getting β_1

$\beta_1 = \dim(\ker(\partial_1)) - \operatorname{rank}(\partial_2))$
$\dim(\ker(\partial_1)) = 4$
$\operatorname{rank}(\partial_2) = 3$
$\beta_1 = 1$



Limitations

Bottleneck

Enumerating cliques.
- More cliques $\rightarrow$ exponentially more computation

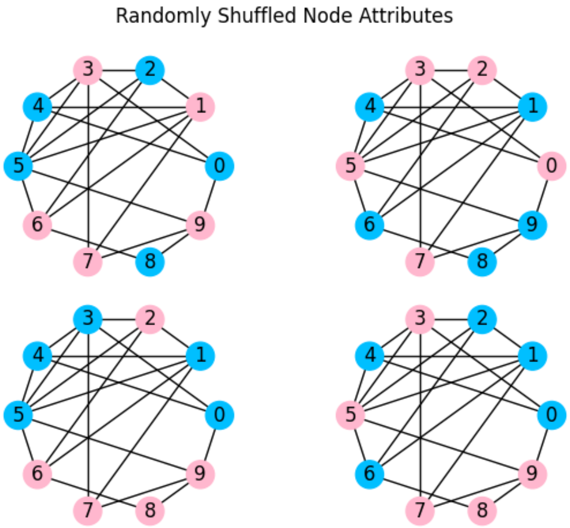
- Things to consider:
  - Graph size
  - Density
  - # of colors
  - # of permutations
- Massive networks are impossible to compute

Hypotheses

H_0 : The graph's homology is independent of the vertex labeling.

H_a : The graph's homology is not independent of the vertex labeling.

Randomly Shuffle Node Labels



Computing a Test Statistic

Mahalanobis Distance ($D^2(\mathbf{x})$)

$$D^2(\mathbf{x}) = (\mathbf{x} - \mu)^T \mathsf{Cov}^{-1} (\mathbf{x} - \mu),$$

where μ is the mean vector.

Permutation Distribution of The Test Statistic

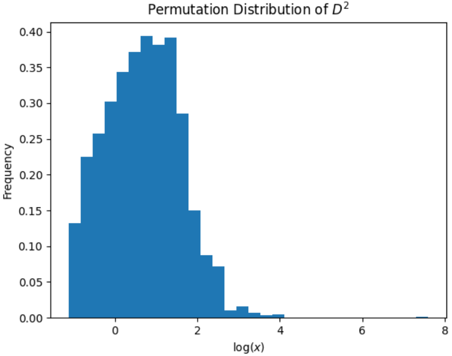


Figure: Permutation distribution of the homology of the C. elegans neural network.

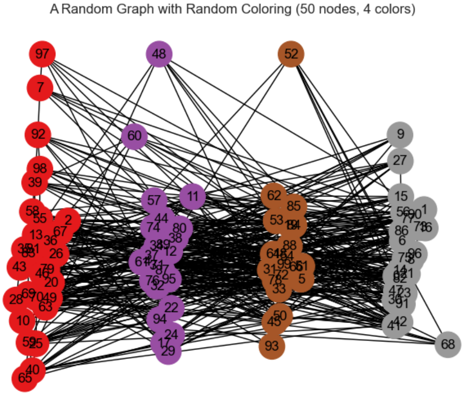
Is our Result Significant?

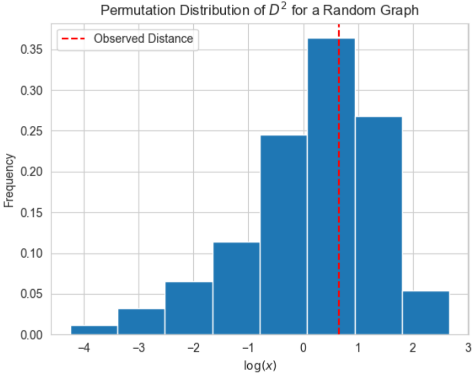
Choose a significance level α. Compute

$$p = \frac{\sum \left(D^2(\mathbf{x}) \geq D^2(\mathbf{x}_{\text{observed}}) \right)}{n}$$

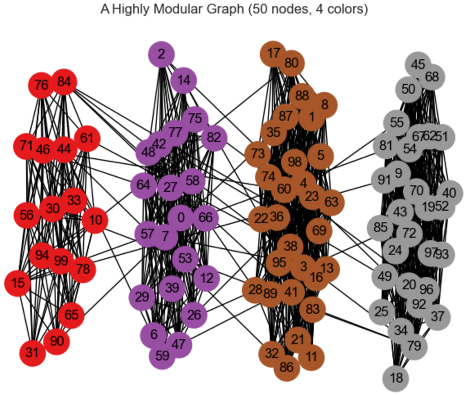
1. $p < \alpha$: reject H_0.
2. $p \geq \alpha$: do *not* reject H_0.

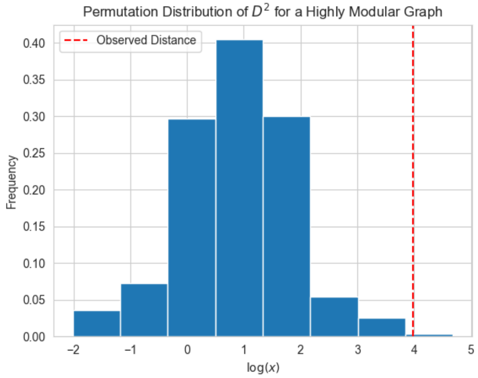
The Null Model: Example





The Alternative Model: Example





Thank you!