motifs
graphlets
network alignment

SYSM 6302

CLASS 17

Motifs - small subgraphs with particular significant structures/patterns

LOCAL $\overset{?}{\underset{?}{\text{OR}}}$ GLOBAL UNDIRECTED $\overset{?}{\underset{?}{\text{OR}}}$ DIRECTED INDUCED $\overset{?}{\underset{?}{\text{OR}}}$ NOT-INDUCED

Comparative Baseline Randomized Graph, maintaining:

1.

2.

Significant structures arise due to:

1.

2.

Motifs - small subgraphs with particular significant structures/patterns

$P \leq 0.01$

LOCAL ? OR ? GLOBAL

3/4-node patterns

UNDIRECTED ? OR ? DIRECTED

could be applied to undirected also

INDUCED ? OR ? NOT-INDUCED

can include additional edges in full network

Comparative Baseline Randomized Graph, maintaining:

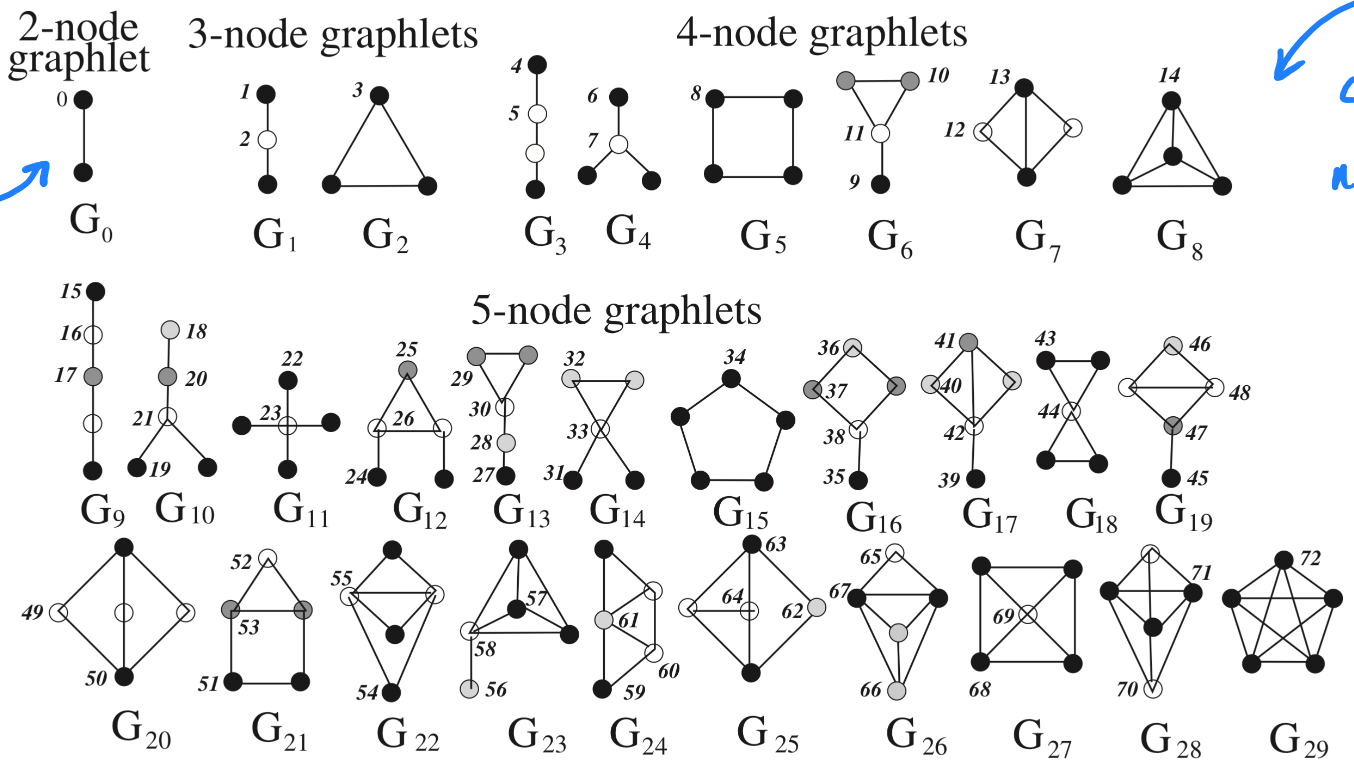
1. single-node degree characteristics
2. same number of n-1 subgraphs

Significant structures arise due to:

1. Different network functions require different structures (i.e., energy flow versus information processing)
2. Network evolution processes may allow for some structures and not others

Graphlets – small induced subgraphs that generalize the degree distribution

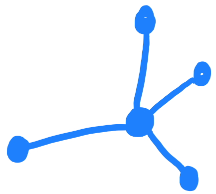
Degree distribution counts the number of 2-node graphlets



Some graphlets admit automorphisms, meaning some nodes are indistinguishable

They call the "different" nodes as "orbits"

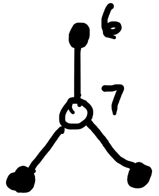
→ 72 2/3/4/5 orbits



degree 4 since node touches 4 G_0 graphlets

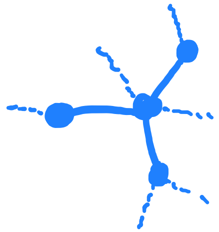
Graphlet degree distribution is the record of how many nodes touch how many graphlets (as a particular orbit)

G_4:

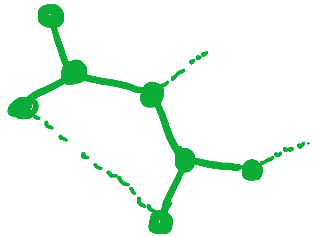


graphlet deg. dist
for orbit 6

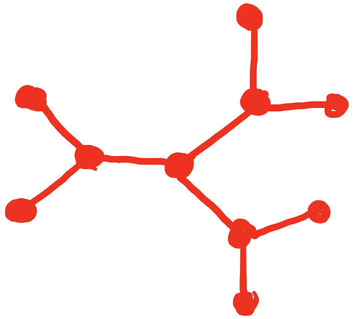
\# of nodes that
touch one G_4 as orbit 6



\# of nodes that
touch two G_4 as orbit 6



\# of nodes that
touch three G_4 at orbit 6



⇒ 72 graphlet degree distributions

↳ These distributions are clearly not independent

↳ Comparing networks requires normalizing and aggregating these 72 distributions

Network Alignment

→ Attempts to find a rough mapping between networks when an isomorphism does not exist

→ Local approaches map subgraphs and typically they can overlap
↳ leads to many-to-many type mappings
↳ typically an easier problem

→ Global maps entire network (all nodes)

"similarity" can be determined by:
① similarity - good match if nbrs are good matches
regular equivalence
② graphlets - good match if extended neighborhood is topologically similar comparing graphlet counts

① Compute "similarity" scores between all nodes in G_1 and all nodes in G_2.

② Greedily or optimally select mapping that connects most similar nodes.

Evaluate Alignment solution: $G_1 = (V_1, E_1)$ $G_2 = (V_2, E_2)$ $f: V_1 \to V_2$

$\to f(V_1) = \tilde{V}_2 \subseteq V_2$ ← always assume $|V_1| \leq |V_2|$

$\to \tilde{G}_2 = (\tilde{V}_2, \tilde{E}_2)$ ← induced subgraph $\tilde{E}_2 = \{(u,v) \mid (u,v) \in E_2 ; u,v \in f(V_1)\}$

Edge Correctness: $\dfrac{|f(E_1)|}{|E_1|}$ $f(E_1) = \{(f(u), f(v)) \in E_2 \mid (u,v) \in E_1\}$ # of mapped edges

what about missed edges in E_2?

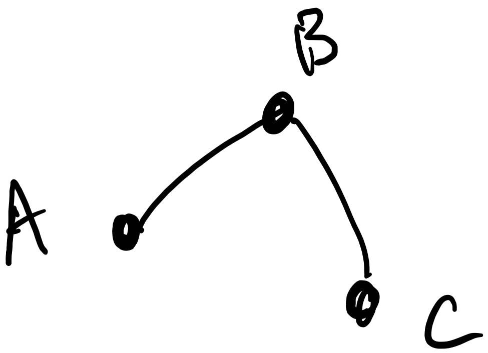
$|f(E_1)| \leq |E_1|$

Induced Conserved Structure: $\dfrac{|f(E_1)|}{|\tilde{E}_2|}$ $|f(E_1)| \leq |\tilde{E}_2|$ Now not penalizing for missing edges in E_1

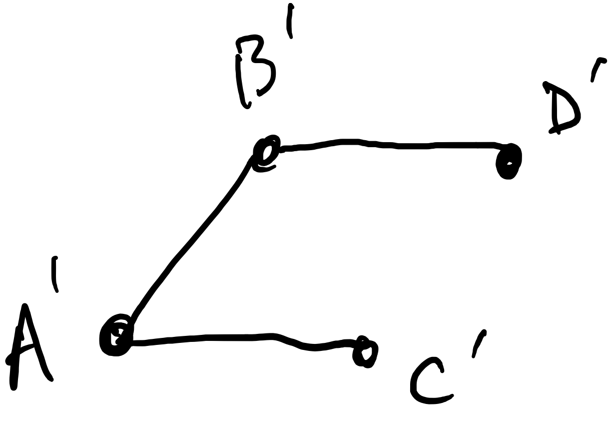
Symmetric Structure Score: $\dfrac{|f(E_1)|}{|E_1| + |\tilde{E}_2| - |f(E_1)|}$

Black-box optimization techniques used to do network alignment by directly optimizing the edge-based alignment scores

G_1



G_2

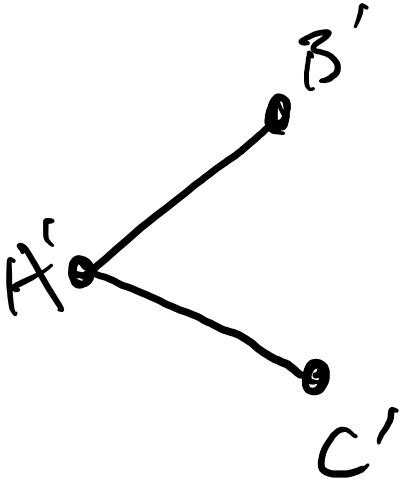


$|E_2| = 3$

$|E_1| = 2$

$f(E_1) = \{(A', B')\}$

$\tilde{G}_2$



$|\tilde{E}_2| = 2$