# subgraphs or *fragments*

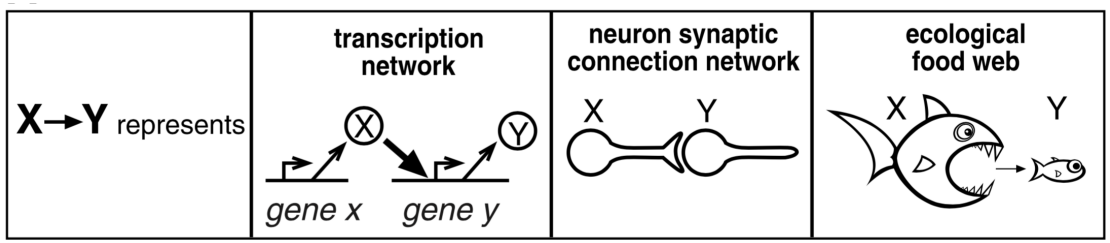
introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2024/25

# fragments *definition*

— small *subgraphs* are *building blocks* of networks

— *subgraphs* characterize *local network structure*




— *fragments* = *connected subgraphs* of networks [EK15]

— *motifs* = *frequent non-induced* subgraphs [MSOI+02]

— *graphlets* = *specific induced* subgraphs [PCJ04]

see mfinder and orca for implementations

# network *motifs*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2024/25

# motifs *definition*

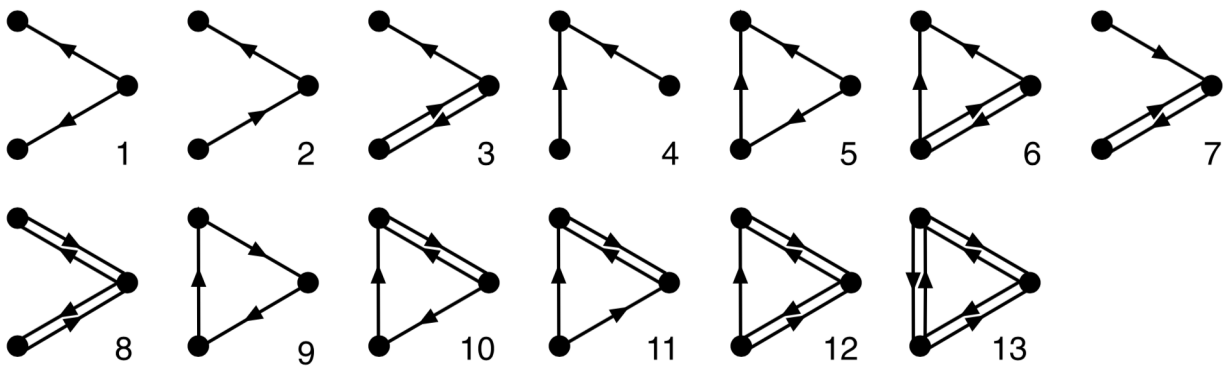
— *fragments* characterize *network-wise local structure*

— *motifs* are *frequent non-induced fragments* [MSOI+02]

probability of *motif appearing in random graph*

*equal or greater number of times* is $< 0.01$

— (*un*)*directed motifs* consisting of *three to five/six/seven nodes*



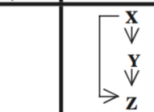
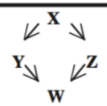
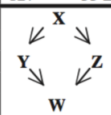
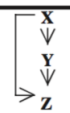
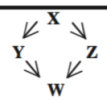
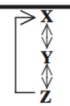
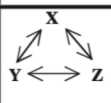
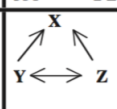

all 13 directed three-node motifs

# motifs *significance*

— *motif significance* $Z$ with *normal distribution* $N(0,1)$
  - $\widetilde{n}_i$ is *number of motifs* $i$ in *random graph* with *variance* $\widetilde{\sigma}_i^2$
  - $n_i$ is *number of motifs* $i$ in *real network*

$$Z_i = \frac{n_i - \langle \widetilde{n}_i \rangle}{\widetilde{\sigma}_i} \qquad n_i - \langle \widetilde{n}_i \rangle > 0.1 \langle \widetilde{n}_i \rangle$$

— $\widetilde{n}/\widetilde{\sigma}$ estimated by *motif preserving randomization* [MSOI+02]

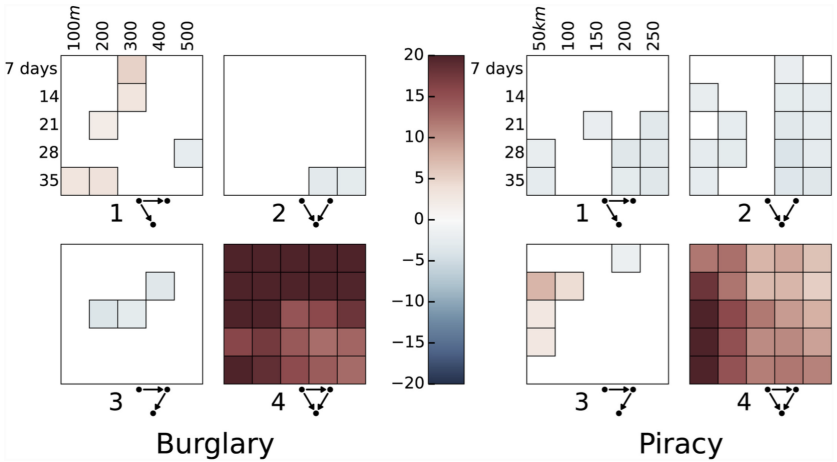
| Network | Nodes | Edges | $N_{real}$ | $N_{rand} \pm SD$ | Z score | $N_{real}$ | $N_{rand} \pm SD$ | Z score | $N_{real}$ | $N_{rand} \pm SD$ | Z score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Neurons** | | |  | | **Feed-forward loop** |  | | **Bi-fan** |  | | **Bi-parallel** |
| *C. elegans*† | 252 | 509 | 125 | 90 ± 10 | 3.7 | 127 | 55 ± 13 | 5.3 | 227 | 35 ± 10 | 20 |
| **Food webs** | | |  | | **Three chain** |  | | **Bi-parallel** | | | |
| Little Rock | 92 | 984 | 3219 | 3120 ± 50 | 2.1 | 7295 | 2220 ± 210 | 25 | | | |
| Ythan | 83 | 391 | 1182 | 1020 ± 20 | 7.2 | 1357 | 230 ± 50 | 23 | | | |
| St. Martin | 42 | 205 | 469 | 450 ± 10 | NS | 382 | 130 ± 20 | 12 | | | |
| **Electronic circuits (forward logic chips)** | | |  | | **Feed-forward loop** |  | | **Bi-fan** |  | | **Bi-parallel** |
| s15850 | 10,383 | 14,240 | 424 | 2 ± 2 | 285 | 1040 | 1 ± 1 | 1200 | 480 | 2 ± 1 | 335 |
| s38584 | 20,717 | 34,204 | 413 | 10 ± 3 | 120 | 1739 | 6 ± 2 | 800 | 711 | 9 ± 2 | 320 |
| s38417 | 23,843 | 33,661 | 612 | 3 ± 2 | 400 | 2404 | 1 ± 1 | 2550 | 531 | 2 ± 2 | 340 |
| **World Wide Web** | | |  | | **Feedback with two mutual dyads** |  | | **Fully connected triad** |  | | **Uplinked mutual dyad** |
| nd.edu§ | 325,729 | 1.46e6 | 1.1e5 | 2e3 ± 1e2 | 800 | 6.8e6 | 5e4±4e2 | 15,000 | 1.2e6 | 1e4 ± 2e2 | 5000 |

# motifs *examples*

## *motif Z-scores* of class *software networks* [VS05]

| Network | Nodes | Edges | $N_{real}$ | $N_{rand}$ | $Z_{score}$ | $N_{real}$ | $N_{rand}$ | $Z_{score}$ | $N_{real}$ | $N_{rand}$ | $Z_{score}$ | $N_{real}$ | $N_{rand}$ | $Z_{score}$ | $N_{real}$ | $N_{rand}$ | $Z_{score}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Software Networks (medium)** | | |  FFL S38 | | |  Bi-parallel S904 | | |  FFL extended S344 | | | | | | | | |
| Fairuse | 106 | 180 | 41 | 11.6±3.3 | 8.94 | 18 | 7.3±3.5 | 3.01 | 33 | 9.5±5.5 | 4.25 | | | | | | |
| Aime | 143 | 319 | | n/a | | 230 | 30.8±19 | 10.46 | 55 | 31.8±9.3 | 2.49 | | | | | | |
| Filezilla221a | 186 | 331 | 77 | 29.4±6.1 | 7.86 | 30 | 10.2±4.5 | 4.38 | | n/a | | | | | | | |
| Aztec | 255 | 391 | 68 | 26±5.4 | 7.82 | 25 | 10.6±4.6 | 3.15 | 68 | 14±5.9 | 9.08 | | | | | | |
| Exult | 261 | 504 | 107 | 56.3±8.4 | 6.01 | 86 | 30.2±8 | 6.75 | 182 | 80.2±19.6 | 5.18 | | | | | | |
| **Software Networks (large)** | | |  Bi-fan S204 | | |  Bi-parallel S904 | | |  Multi-X S2252 | | |  Multi-Z S206 | | |  Multi-Y S2190 | | |
| blender226 | 495 | 834 | 486 | 138±30.3 | 11.4 | 33 | 16±5.2 | 3.2 | 123 | 7.8±5.8 | 20 | 22 | 3.7±3.6 | 5.04 | 18 | 4.2±3.2 | 4.37 |
| gtk221 | 748 | 1147 | 644 | 189±33.8 | 13.4 | 119 | 25±7.5 | 12.5 | 173 | 26±16.7 | 8.8 | 21 | 2.0±1.7 | 10.9 | 19 | 4.2±2.7 | 5.41 |
| vtk | 771 | 1362 | 512 | 262±39.9 | 6.27 | 295 | 69±14.3 | 15.8 | 41 | 6.3±3.2 | 10.7 | 193 | 27.7±14.7 | 11.2 | 122 | 12.8±5.8 | 18.7 |
| java2 | 1364 | 1947 | 816 | 189±35.5 | 17.7 | 173 | 48±10.8 | 11.5 | 345 | 18.4±14 | 23.3 | 22 | 2.2±2.1 | 9.5 | 17 | 3.8±2.2 | 5.99 |
| prorally | 1993 | 4987 | 22750 | 1840±171 | 122 | 3848 | 322±34.2 | 103.1 | 1080 | 144±50.6 | 18.4 | 210 | 28.7±9.2 | 19.8 | 1318 | 55.5±14.7 | 85.9 |
| **Software Networks (large)** | | |  FFL S38 | | |  S472 | | |  Multi-X incomp. S2186 | | |  Multi-X incomp S408 | | |  FFL extended S344 | | |
| blender226 | | | 126 | 33.8±6.1 | 15 | | N/A | | 1976 | 766±162 | 7.43 | 436 | 196±65 | 3.7 | 94 | 26.2±8.6 | 7.88 |
| gtk221 | | | 118 | 47.7±9.6 | 7.31 | 15 | 3.1±2.3 | 5.06 | 4177 | 1941±398 | 5.6 | 1462 | 748±261 | 2.73 | 188 | 68.1±13.7 | 8.73 |
| vtk | | | 229 | 81.6±10.6 | 13.9 | 30 | 13±6.7 | 2.53 | 707 | 388±61.6 | 5.17 | 333 | 217±44 | 2.62 | 718 | 212.1±49.1 | 10.3 |
| java2 | | | 176 | 46.2±9 | 14.4 | 8 | 1.8±1.4 | 4.57 | 10212 | 6346±1180 | 3.2 | 2494 | 1397±522 | 2.1 | 257 | 52.5±17.6 | 11.6 |
| prorally | | | 1169 | 272±21 | 42.6 | 282 | 30.8±10.4 | 24.2 | 25099 | 15482±1750 | 5.5 | 5742 | 4163±752 | 2.1 | 2909 | 736±101.4 | 21.4 |

## *motif Z-scores* of spatio-temporal *crime networks* [DM15]



# motifs *profiles*

— *motif significance profile SP* [MSOI+02] defined as
  – $Z_i$ is *significance of motif i* in *real network*

$$SP_i = \frac{Z_i}{\sqrt{\sum_i Z_i^2}} \qquad Z_i = \frac{n_i - \langle \widetilde{n}_i \rangle}{\widetilde{\sigma}_i} \qquad n_i \geq 4$$

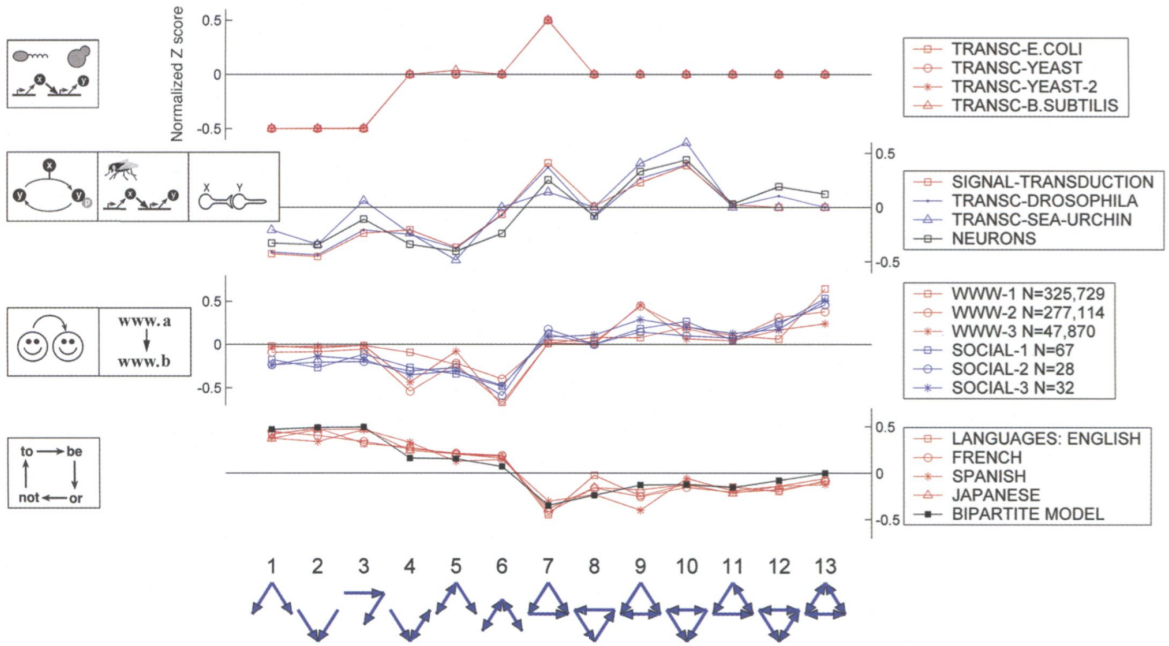
— *motif abundance/ratio profile RP* [MIK+04] defined as
  – $A_i$ is *abundance of motif i* in *real network*

$$RP_i = \frac{A_i}{\sqrt{\sum_i A_i^2}} \qquad A_i = \frac{n_i - \langle \widetilde{n}_i \rangle}{n_i + \langle \widetilde{n}_i \rangle + \epsilon_i} \qquad \epsilon_i = 4$$

# motifs *families*

— directed *motif significance profiles* [MSOI$^+$02]

— profiles reveal (*super*)*families of real networks*




all 13 directed three-node motifs

# network *graphlets*

introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2024/25

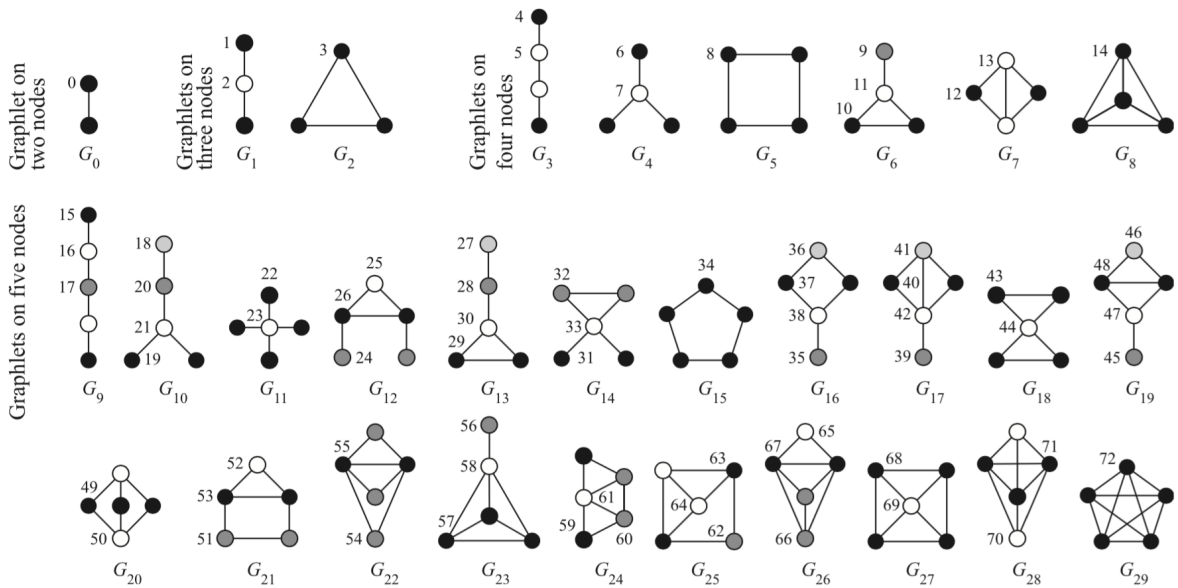
# graphlets *definition*

— *fragments* characterize *node-wise local structure*

— *graphlets* are *specific induced fragments* [PCJ04]

— *graphlet orbits* are *automorphisms of graphlets* [Pržˇ07]

— (*un*)*directed graphlets* consisting of *three to five/. . . nodes*



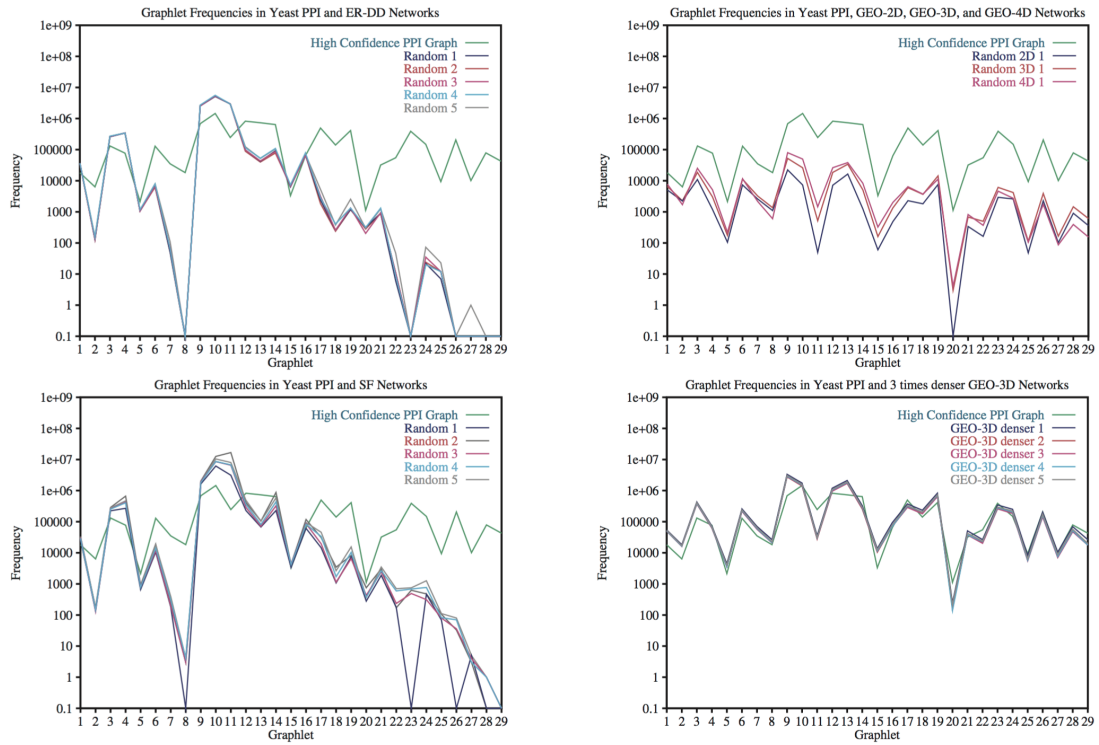

all 30 undirected two- to five-node graphlets with 73 orbits

# graphlets *frequency*

— *relative graphlet frequency* $F$ [PCJ04] defined as

– $n_i$ is *number of graphlets* $i$ in *real network*

$$F_i = \frac{n_i}{\sum_i n_i}$$

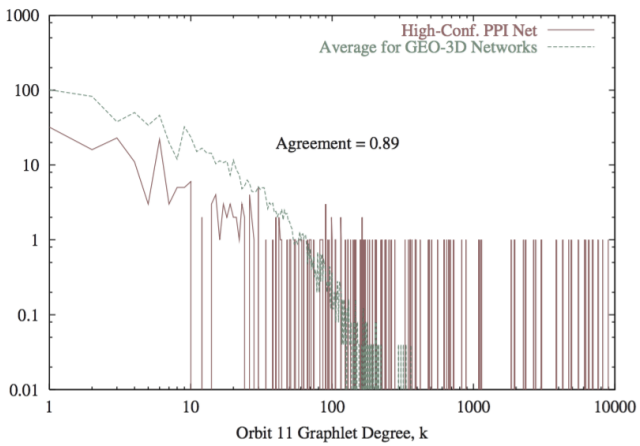


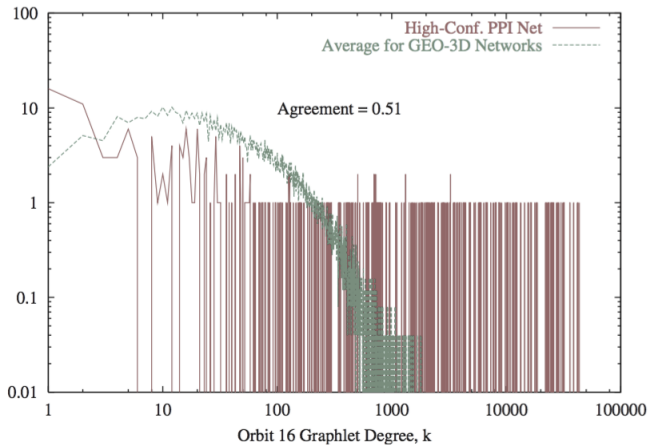

graphlet frequency in protein network and random graphs

# graphlets *distribution*

— *i-th orbit graphlet degree distribution* $p_k^i$ [Pržˇ07] defined as
  - $p_k^0$ is *degree distribution* $p_k$ of *real network*
  - $p_k^i$ is *graphlet degree distribution* for *i-th orbit*
  - $\widetilde{p}_k^i$ is *scaled graphlet degree distribution* for *i-th orbit*

$$\widetilde{p}_k^i \sim p_k^i / k \qquad \widetilde{p}_k^0 = p_k^0 = p_k$$




11th and 16th orbit graphlet degree distributions of protein network and random graph

# graphlets *agreement*

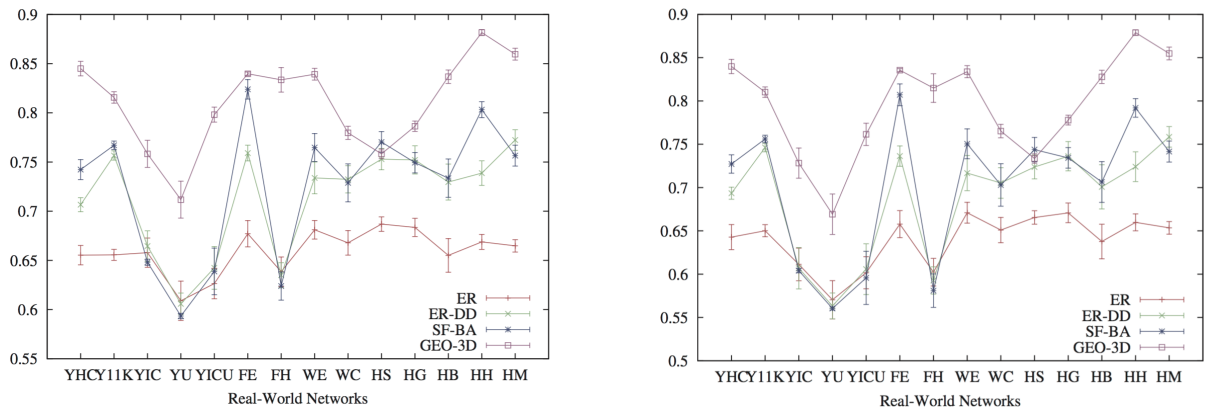
— *i-th orbit graphlet agreement* $A_i$ [Pržž07] defined as

- $\widetilde{p}_k^i$ is *i-th orbit graphlet degree distribution* of *first network*
- $\widetilde{q}_k^i$ is *i-th orbit graphlet degree distribution* of *second network*

$$A_i = 1 - \sqrt{\frac{1}{2}\sum_k \left(\log \widetilde{q}_k^i - \log \widetilde{p}_k^i\right)^2}$$

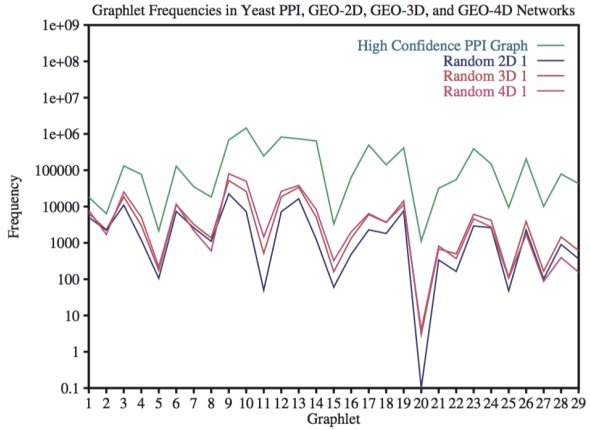
— *arithmetic/geometric graphlet agreement* $A$ defined as

$$A = \frac{1}{73}\sum_i A_i \qquad A = \left(\prod_i A_i\right)^{\frac{1}{73}}$$

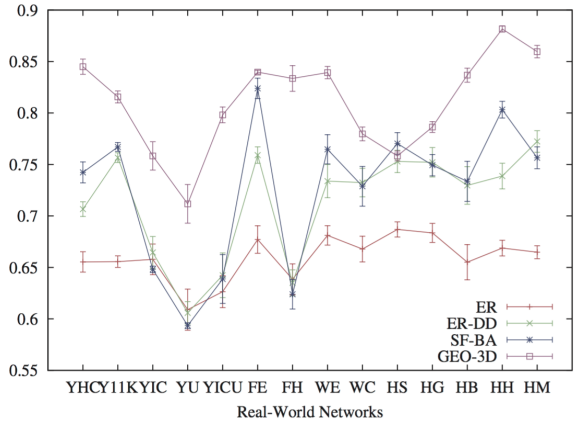



arithmetic/geometric graphlet agreement of protein networks and random graphs

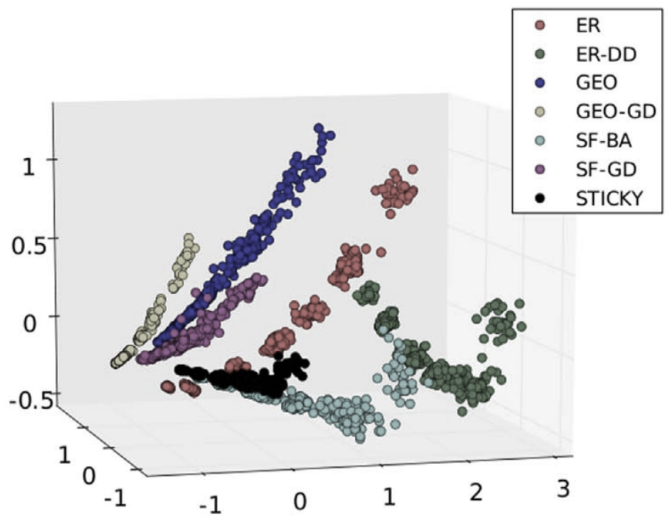
# graphlets *measures*

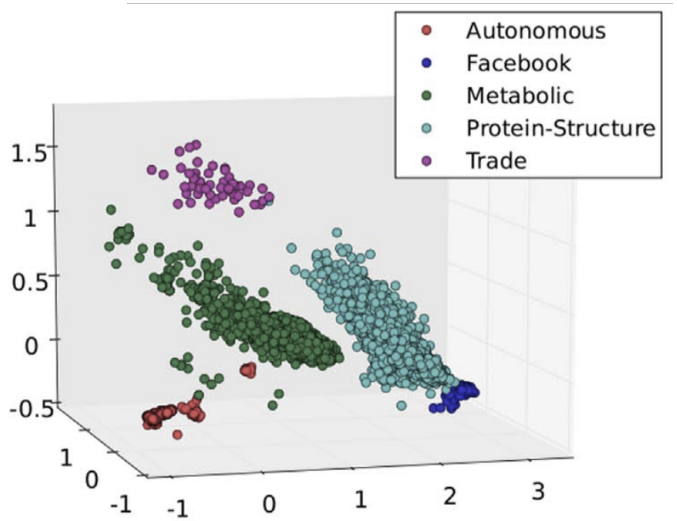



*relative graphlet frequency* [PCJ04]




*graphlet distribution agreement* [Pržž07]




*graphlet correlation matrix* and *distance* [YMDD+14]

# fragments *references*


A.-L. Barabási.
*Network Science*.
Cambridge University Press, Cambridge, 2016.

Toby Davies and Elio Marchione.
Event networks and the identification of crime pattern motifs.
*PLoS ONE*, 10(11):e0143638, 2015.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
*Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition*.
Cambridge University Press, Cambridge, 2011.

David Easley and Jon Kleinberg.
*Networks, Crowds, and Markets: Reasoning About a Highly Connected World*.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
*A First Course in Network Theory*.
Oxford University Press, 2015.

Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon.
Superfamilies of evolved and designed networks.
*Science*, 303(5663):1538–1542, 2004.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon.
Network motifs: Simple building blocks of complex networks.
*Science*, 298(5594):824–827, 2002.

# fragments *references*


Mark E. J. Newman.
*Networks*.
Oxford University Press, Oxford, 2nd edition, 2018.

N. Pržulj, D. G. Corneil, and I. Jurisica.
Modeling interactome: Scale-free or geometric?
*Bioinformatics*, 20(18):3508–3515, 2004.

Nataša Pržulj.
Biological network comparison using graphlet degree distribution.
*Bioinformatics*, 23(2):e177–e183, 2007.

Sergi Valverde and Ricard V. Solé.
Network motifs in computational graphs: A case study in software architecture.
*Phys. Rev. E*, 72(2):026107, 2005.

Ömer Nebil Yaveroğlu, Noël Malod-Dognin, Darren Davis, Zoran Levnajić, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj.
Revealing the hidden language of complex networks.
*Sci. Rep.*, 4:4547, 2014.