



Significance of Directed graphs properties in Complex Systems

Iriarte Delfina (1231682)

Karan Kabbur Hanumanthappa Manjunatha(1236383)



Introduction

- Many networks describing complex systems are directed:
the interactions between elements are not symmetric.
- Directed networks exhibit ***trophic coherence*** q :
there exists a well-defined hierarchy of vertices.
- We will focus on the study of the following aspects related to q :
 - Looplessness
 - Non normality
 - Strong connectivity
- We will use many types of dataset: from genes, metabolites, species, neurons, words, computers, and trading nation

Looplessness in networks is linked to trophic coherence

Feedback is a fundamental process in dynamical systems that occurs when the output of an element is coupled to its input.

Definitions:

Directed configuration ensemble :

set of directed graphs with a given number of nodes and degree sequence.



Basal ensemble:

which is the subset of graphs from the directed configuration ensemble that satisfy the constraint that the proportion of neighbors connected to basal nodes is exactly $k_{in} L_b / L$.

Coherence Ensemble:

ensemble of directed graphs that not only have given in- and out-degree distributions (as in the directed configuration ensemble) but also given trophic coherence

Consider the directed, unweighted graph given by $N \times N$ adjacency matrix A which has L edges:



Branching factor

$$\alpha = \frac{\langle k^{in} k^{out} \rangle}{\langle k \rangle}.$$



Basal node

one with in degree equal to zero



Trophic level

$$s_i = 1 + \frac{1}{k_i^{in}} \sum_j a_{ij} s_j.$$
$$s_i = 1 \ (\forall i \text{ such that } k_i^{in} = 0)$$

Looplessness in networks is linked to trophic coherence

Definitions:



Trophic difference

$$x_{ij} = s_i - s_j.$$



Trophic coherence

$$q = \sqrt{\frac{1}{L} \sum_{ij} a_{ij} x_{ij}^2} - 1.$$

Maximal coherence $q = 0$



Loop exponent

$$\tau = \ln \alpha + \frac{1}{2\tilde{q}^2} - \frac{1}{2q^2},$$

$\tau < 0$ loopless regime
 $\tau > 0$ loopful regime



Basal ensemble

Expectations for the trophic coherence:

$$\tilde{q} = \sqrt{\frac{L}{L_B} - 1}$$

Expectation of branching factor

$$\tilde{\alpha} = \frac{L - L_B}{N - B}$$



Coherence ensemble

$$\overline{\lambda_1} = e^\tau.$$

For all data being used we have remove all self-edges presents

Looplessness in networks is linked to trophic coherence

Directed Acyclic Graphs

Probability that a graph randomly chosen from the coherence ensemble will have exactly m_ν cycles.

$$p(m_\nu) = \binom{\tilde{n}_\nu}{m_\nu} \bar{c}_\nu^{m_\nu} (1 - \bar{c}_\nu)^{\tilde{n}_\nu - m_\nu}$$

Probability that a network from the coherence ensemble would **have no directed cycles of length greater or equal to n**:

$$P_n = \prod_{\nu=n}^{\infty} p(m_\nu = 0).$$

Probability that a network drawn randomly from this ensemble would be acyclic:

$$P_{acyclic} = \prod_{\nu=2}^{\infty} p(m_\nu = 0) = \prod_{\nu=2}^{\infty} \left\{ 1 - \frac{\tilde{\alpha}}{L} \frac{\tilde{q}}{q} \exp \left[\frac{\nu}{2} \left(\frac{1}{\tilde{q}^2} - \frac{1}{q^2} \right) \right] \right\}^{L\alpha^{\nu-1}}.$$

Taking logarithms and considering graphs with sufficiently negative τ :

$$\ln P_{acyclic} \simeq -\frac{\tilde{\alpha} \tilde{q}}{\alpha q} \sum_{\nu=1}^{\infty} e^{\tau \nu};$$

$$P_{acyclic} \simeq \exp \left[-\frac{\tilde{\alpha} \tilde{q}}{\alpha q} \frac{1}{(e^{-\tau} - 1)} \right].$$

Non Normality matrix

An $N \times N$ matrix A is said to be normal if its adjacency matrix commutes with its transpose or conversely it is **non-normal** if $AA^T \neq A^T A$

- A large deviation from normality to indicate a network with a well-defined directionality. This is also what occurs in trophically coherent networks.
- We can quantify a network being highly non-normal by using the Hermici's departure from normality:

$$D_F = \sqrt{\|A\|_F^2 - \sum_i |\lambda_i|^2},$$

In order to compare matrices of different sizes, they use the normalised version:

$$d_F = \frac{D_F}{\|A\|_F}.$$

A normal matrix will have $d_F = 0$, and d_F is closer to 1 the more A departs from normality

Theorem. The expected deviation from normality, $\overline{d_F}$, for digraphs drawn from the coherence ensemble tends to 1 with increasing trophic coherence. That is,

$$\lim_{q \rightarrow 0} \overline{d_F} = 1.$$

Furthermore,

$$\overline{d_F} > \sqrt{1 - \frac{1}{\langle k \rangle}}$$

for digraphs in the $\tau < 0$ regime, where $\langle k \rangle$ is the mean degree.

Coherent networks are highly non normal.

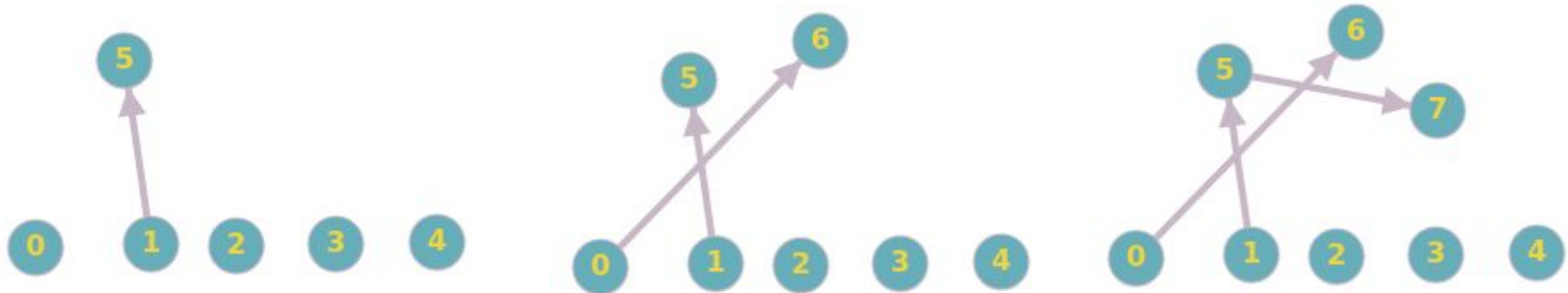
Generalised Preferential-Prey model

Alternative to PPM (networks generated always acyclic).

Possibility to tune trophic coherence.

Algorithm:

- Input: N total nodes; B basal nodes; $\langle k \rangle$ mean degree/ L total number of edges; T temp. (related to incoherence q)
- Begin with **B basal nodes** in directed graph. **Add $(N-B)$ non-basal nodes** one by one. When each node is added **connect it with directed edge** with direction from B to Non basal node at first addition while remaining nodes added have in-connections from both basal and non-basal nodes.. Meaning added node preys on nodes already present.



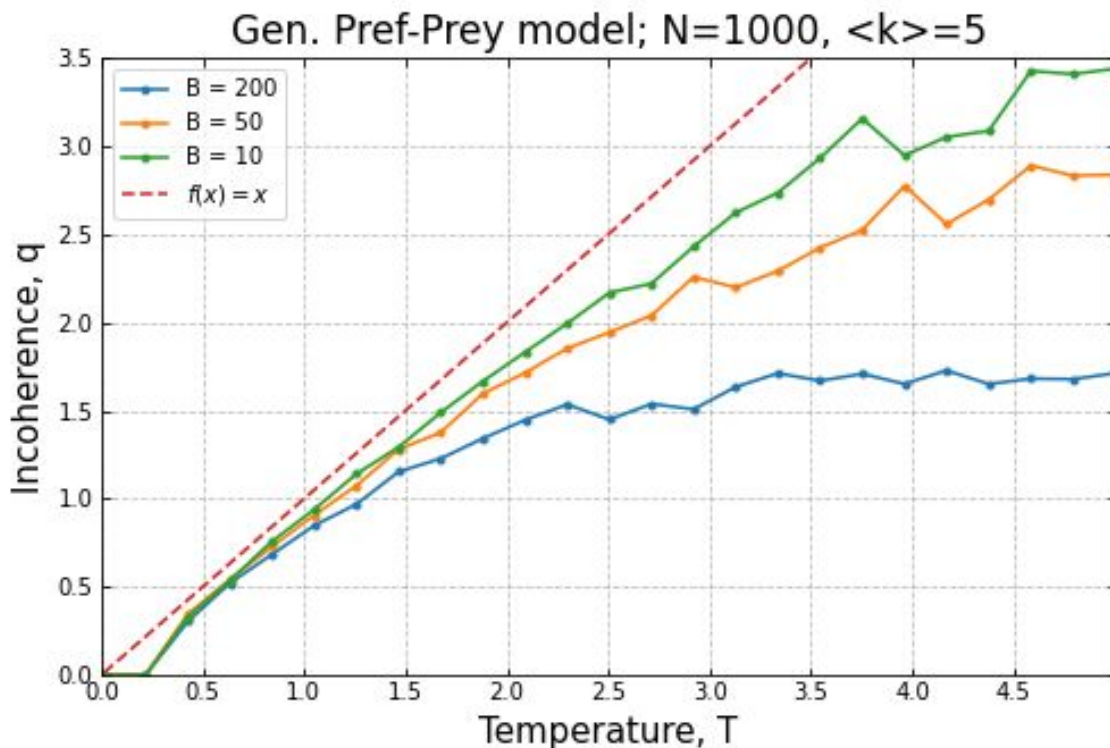
Generalised Preferential-Prey model

Algorithm:

- 1. (N-B) edges are added → (L-N+B) edges are to be added. Put all possible combinations of edges in a list.
- 2. Start a loop to complete L-N+B edges; Compute trophic level 's' of each node. Find trophic difference with i $x_{ij} = s_i - s_j$ always being non basal node and j being either basal or non basal node.
- 3. Choose an edge randomly from list of all possible edges.
- 4. If probability of connection from j to i, $U(0, 1) < P_{ji} = e^{\frac{-(x_{ij}-1)^2}{2T^2}}$ then add edge from j to i and remove that particular edge from list of all possible edges.
- 5. The trophic level values 's' will be updated in the next iteration.
- 6. After all L edges have been added, compute trophic incoherence parameter q of the obtained Di-graph.

"Gaussian distribution of distances x, is found to be a good fit to empirical data on several kinds of networks."

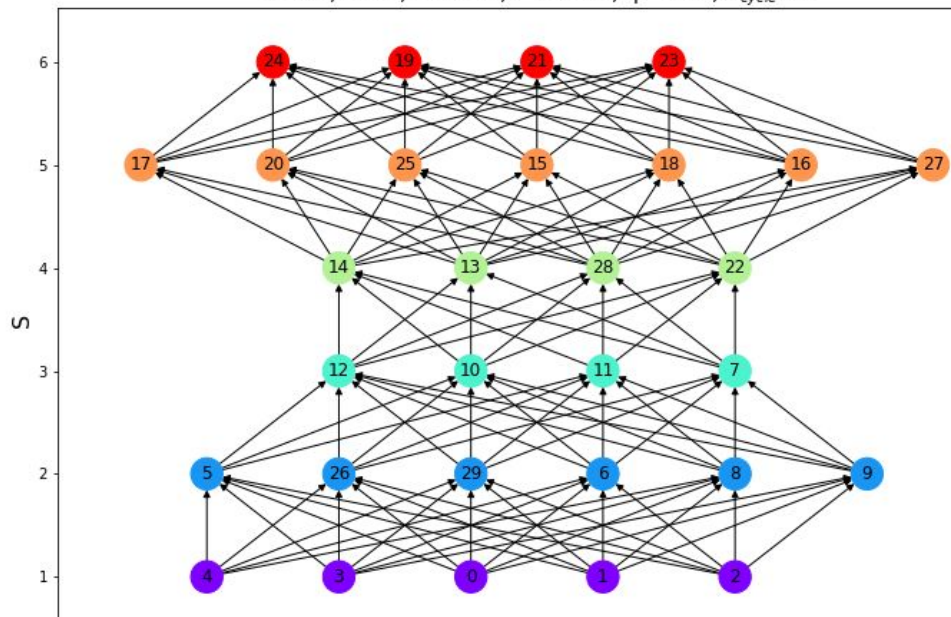
Monotonic relation btw. q and T : Results



- “Temperature” parameter T tunes the degree of trophic coherence.
- $T = 0 \rightarrow$ maximally coherent networks ($q = 0$).
- Incoherence q , increasing monotonically with T .
- Low $T \rightarrow q \approx T$
- High $T \rightarrow$ coherence saturates.

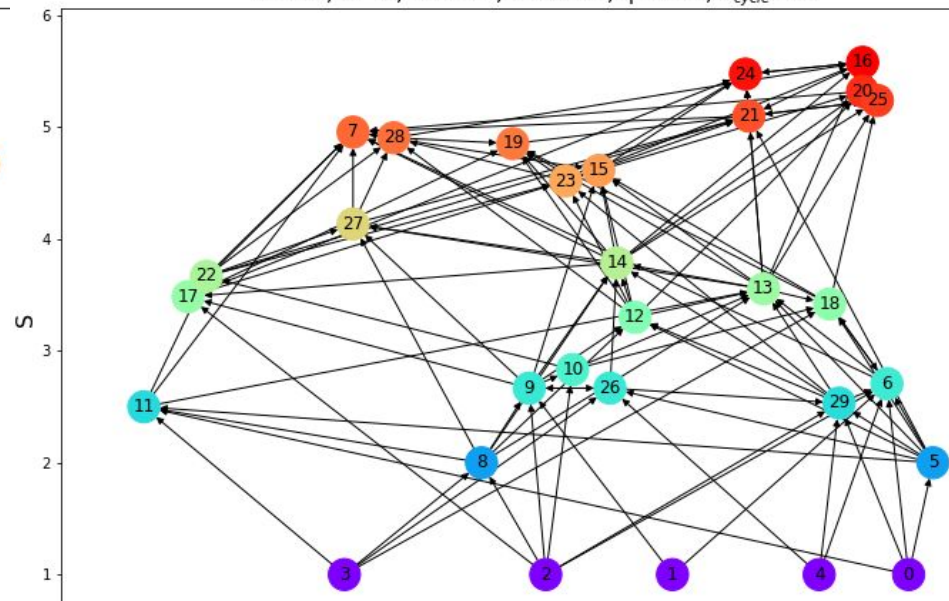
Simulation: Results

Generalised Preferential-Prey model simulation
 $N = 30$; $B = 5$; $\langle k \rangle = 4$; $T = 0.001$; $q = 0.00$; $N_{cycle} = 0$



- Color representation of node:
VIBGYOR; node with higher S values
→ orange/red

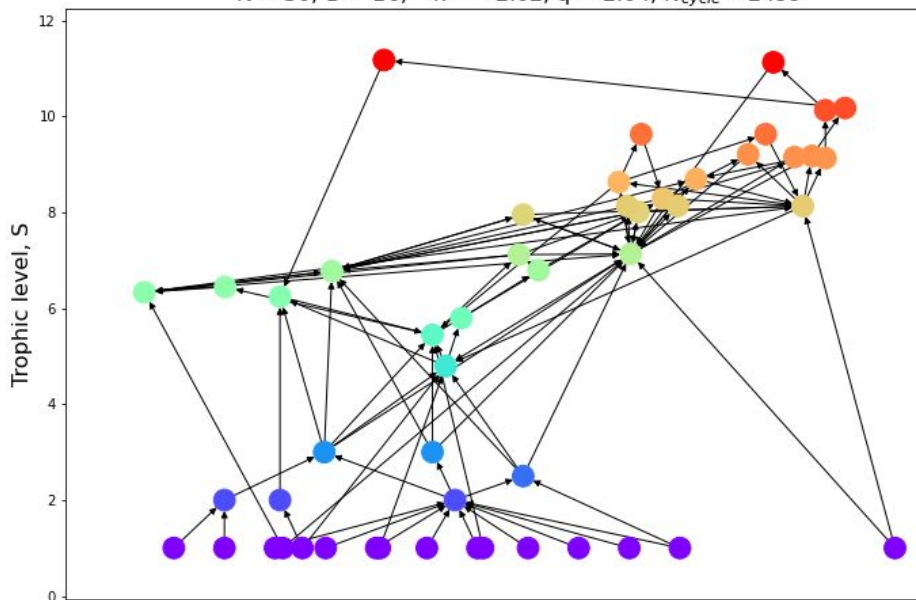
Generalised Preferential-Prey model simulation
 $N = 30$; $B = 5$; $\langle k \rangle = 4$; $T = 1.000$; $q = 0.74$; $N_{cycle} = 31$



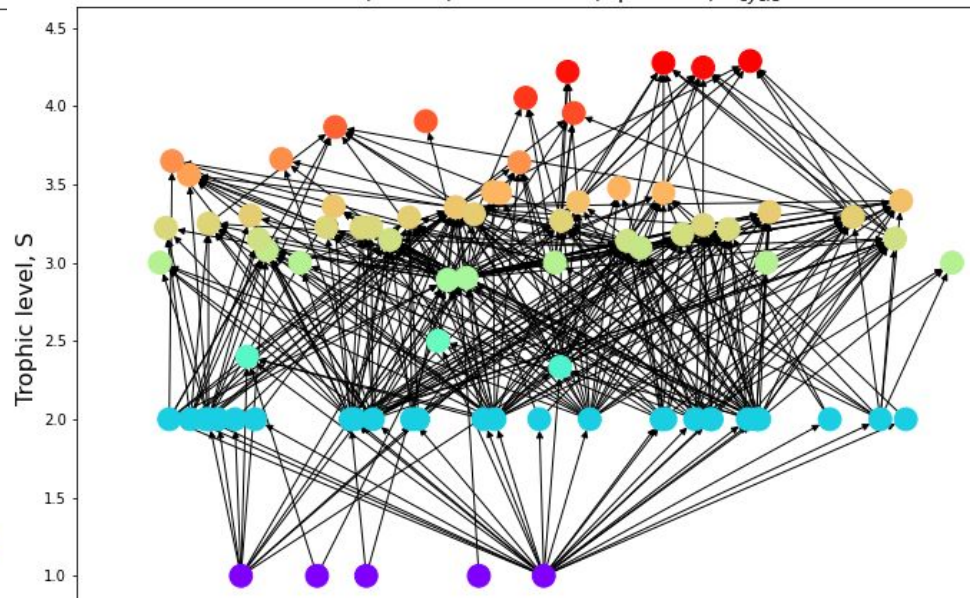
Increase in T → Increase in number of
cycles; highly cyclic

Visualizing Trophic Levels of Real networks

Word network; Green Eggs and Ham, by Dr. Seuss
 $N = 50$; $B = 16$; $\langle k \rangle = 2.02$; $q = 2.04$; $N_{cycle} = 2499$



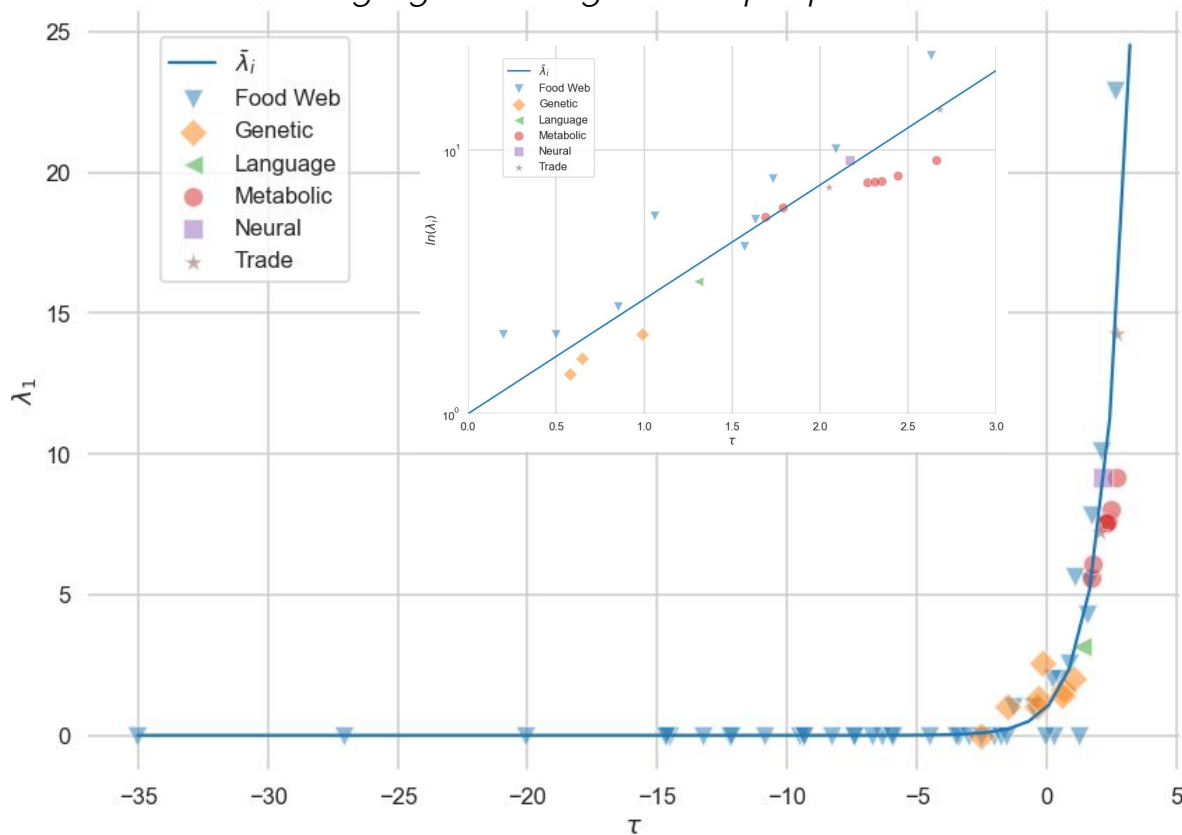
Ythan Estuary food web
 $N = 82$; $B = 5$; $\langle k \rangle = 4.77$; $q = 0.42$; $N_{cycle} = 1$



- If word 1 appears immediately before word 2 in at least one sentence in the text, a directed edge (arrow) is placed from word 1 to word 2.
- The height of each word is proportional to its trophic level.

Looplessness: Results

Leading Eigenvalue against Loop exponent



The coherence ensemble expected value provides a good estimate of almost all of the empirical values.

$$\lambda(\tau < 0) = 0.22 \pm 0.54$$

$$\lambda(\tau > 0) = 6.13 \pm 4.86$$

- Self-loops(cannibalism in food webs) are removed.
- Total 61 networks:
 - $\tau < 0 \rightarrow 36$;
 - $\tau > 0 \rightarrow 25$

Looplessness: Results

	q/q'	s/s'	α / α'	τ	λ_1	$\tau < 0$	Acyclic
Food	0.44 ± 0.17	0.77 ± 0.32	1.01 ± 0.23	-6.23 ± 7.96	1.54 ± 4.04	31/42	74% 31/42
Genetic	0.99 ± 0.05	1.0 ± 0.01	1.19 ± 0.32	-0.34 ± 1.11	1.36 ± 0.71	5/8	63% 1/8
Language	1.01 ± 0.0	1.16 ± 0.0	1.55 ± 0.0	1.31 ± 0.0	3.17 ± 0.0	0/1	0/1
Metabolic	1.81 ± 0.1	1.93 ± 0.12	3.98 ± 0.96	2.22 ± 0.32	7.36 ± 1.11	0/7	0/7
Neural	0.42 ± 0.0	0.39 ± 0.0	1.42 ± 0.0	2.17 ± 0.0	9.15 ± 0.0	0/1	0/1
Trade	1.1 ± 0.04	1.05 ± 0.08	1.19 ± 0.09	2.37 ± 0.32	10.78 ± 3.53	0/2	0/2

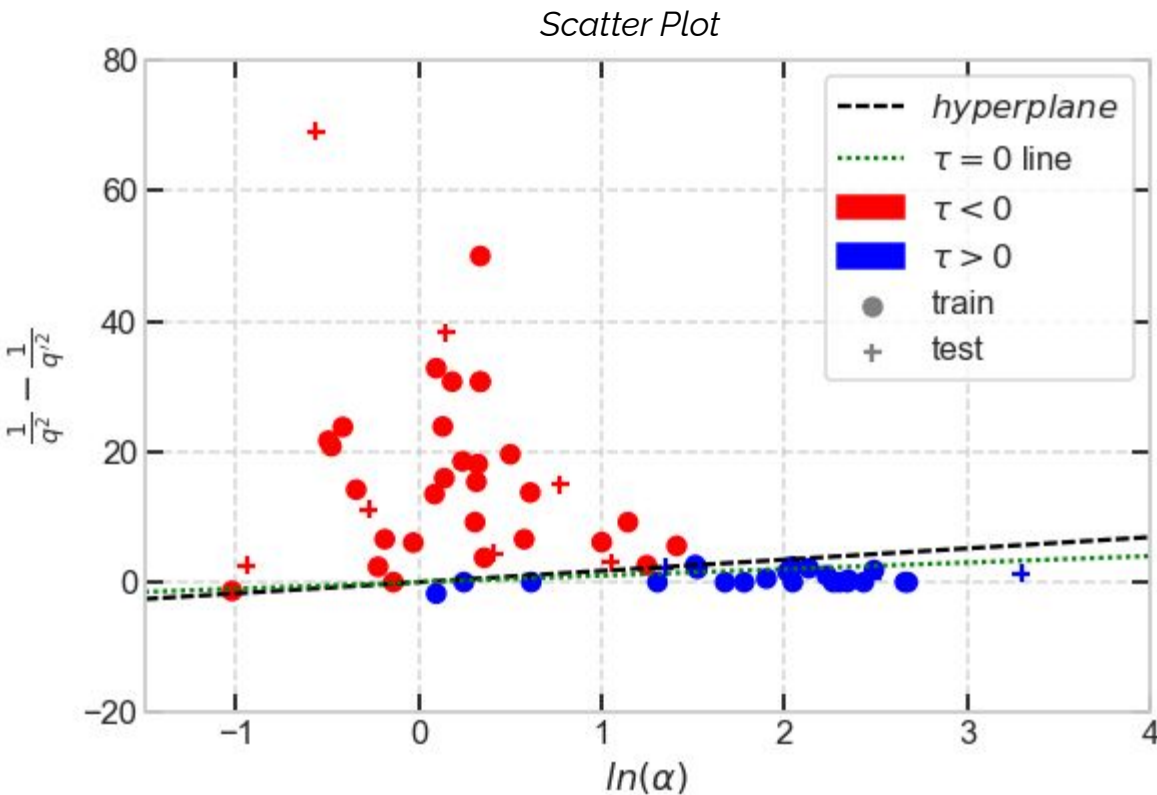
Significantly
coherent

Significantly
Incoherent

Positive correlation btw.
in and out-degrees

Accuracy of τ in
prediction: 87.5%

Looplessness: Results



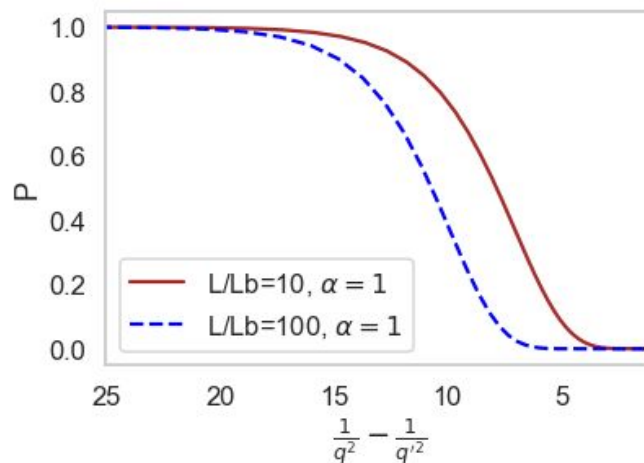
Classification also according to **Perceptron**:

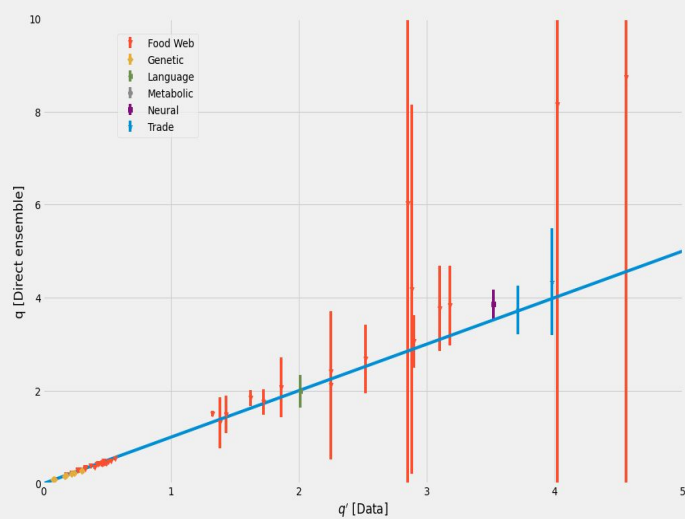
Train-Test split: 85% and 15%

Train accuracy: 100%

Test accuracy: 100%

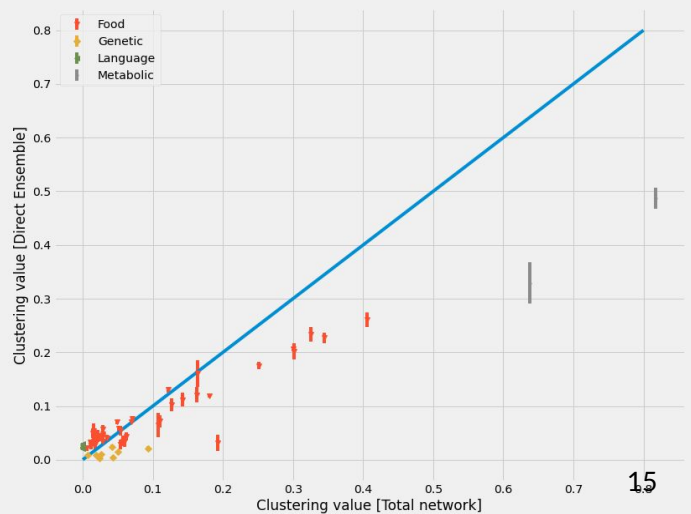
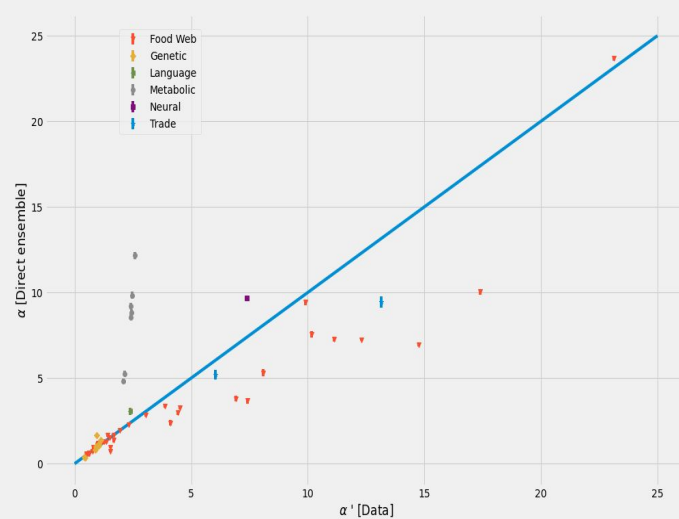
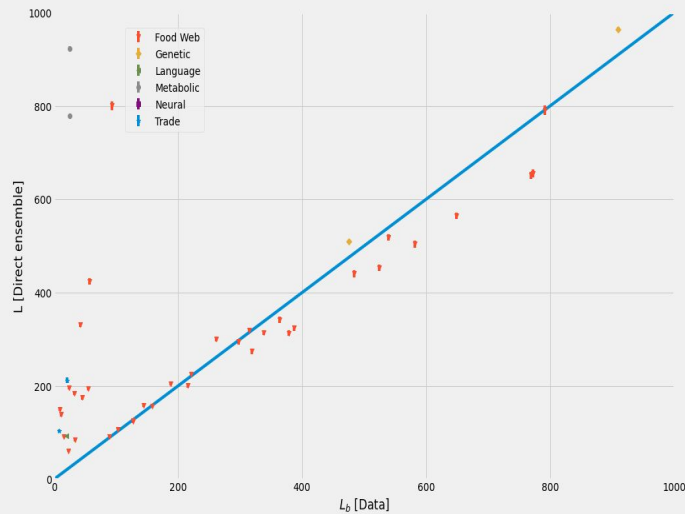
Probability of being acyclic,
 $P(\text{acyclic formula})$

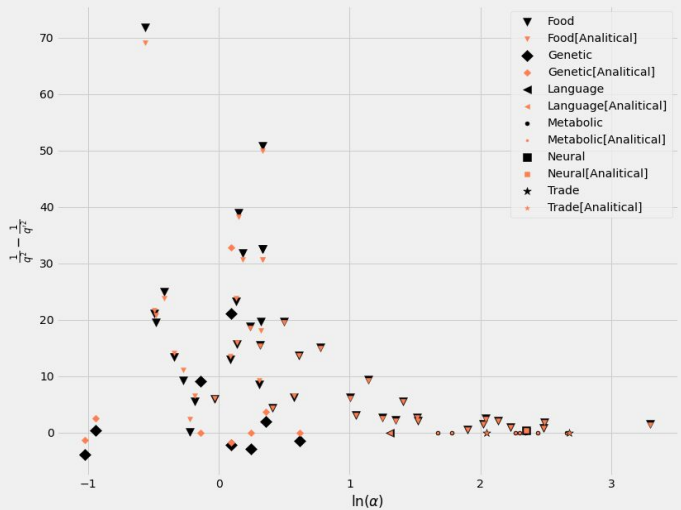
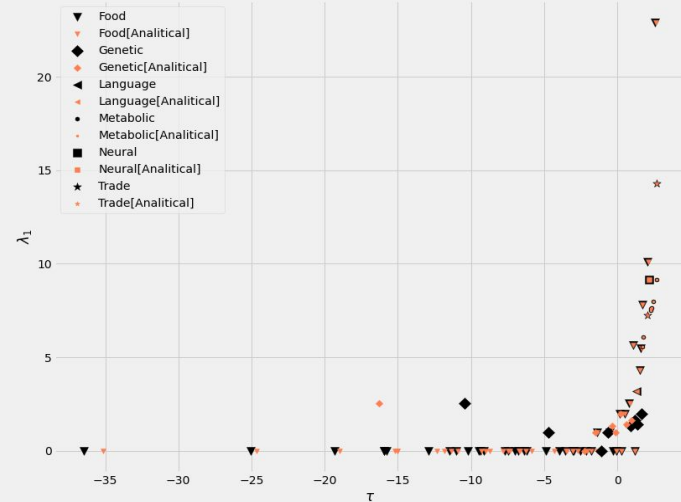




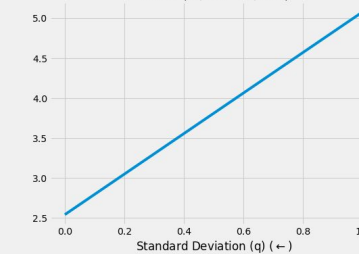
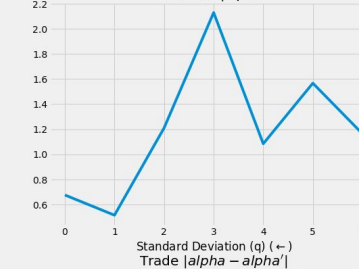
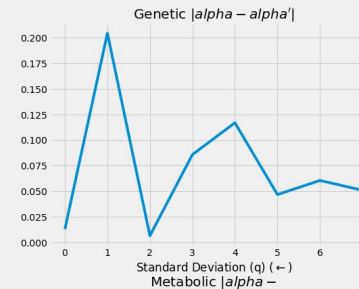
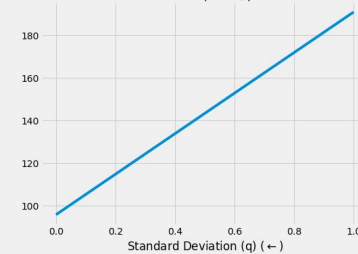
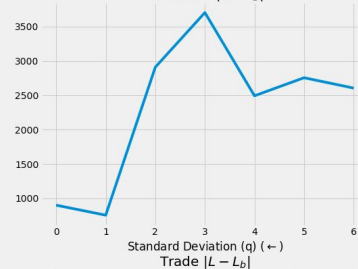
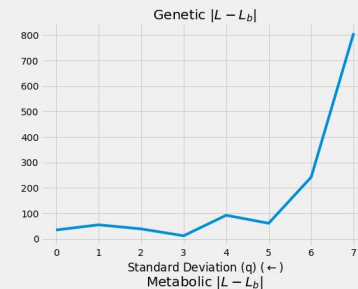
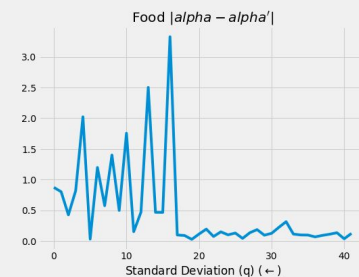
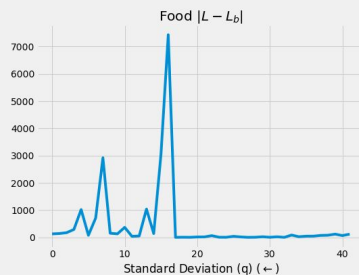
*Main parameters found in the
directed configuration ensemble vs
analytical Basal ensemble*

Comparison among Directed configuration Ensemble and analytical Basal Ensemble



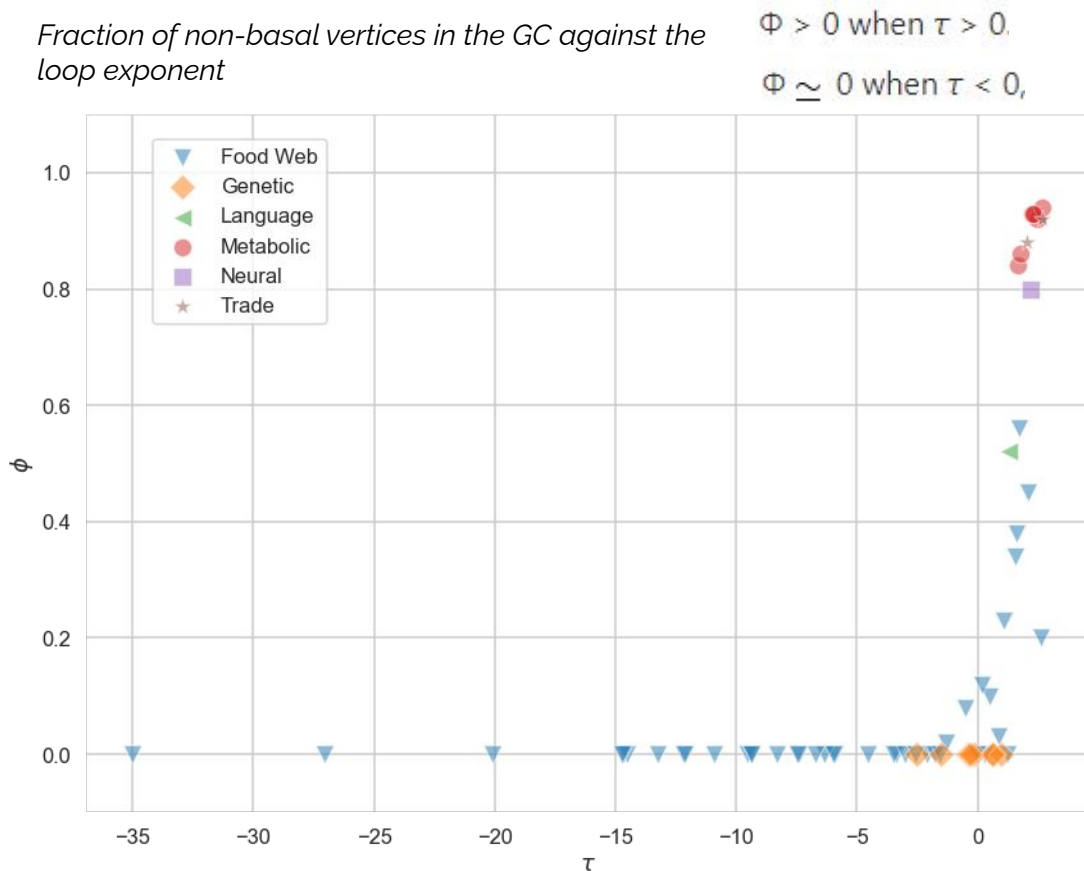


Comparison among ensembles



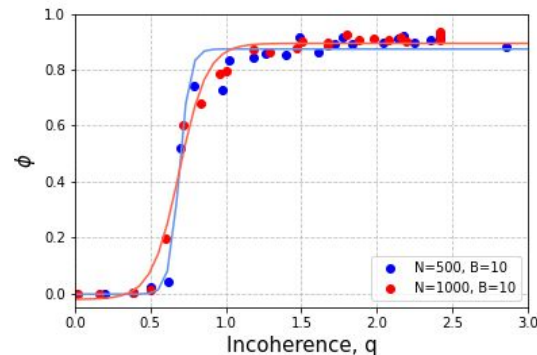
Strong Connectivity: Results

Fraction of non-basal vertices in the GC against the loop exponent



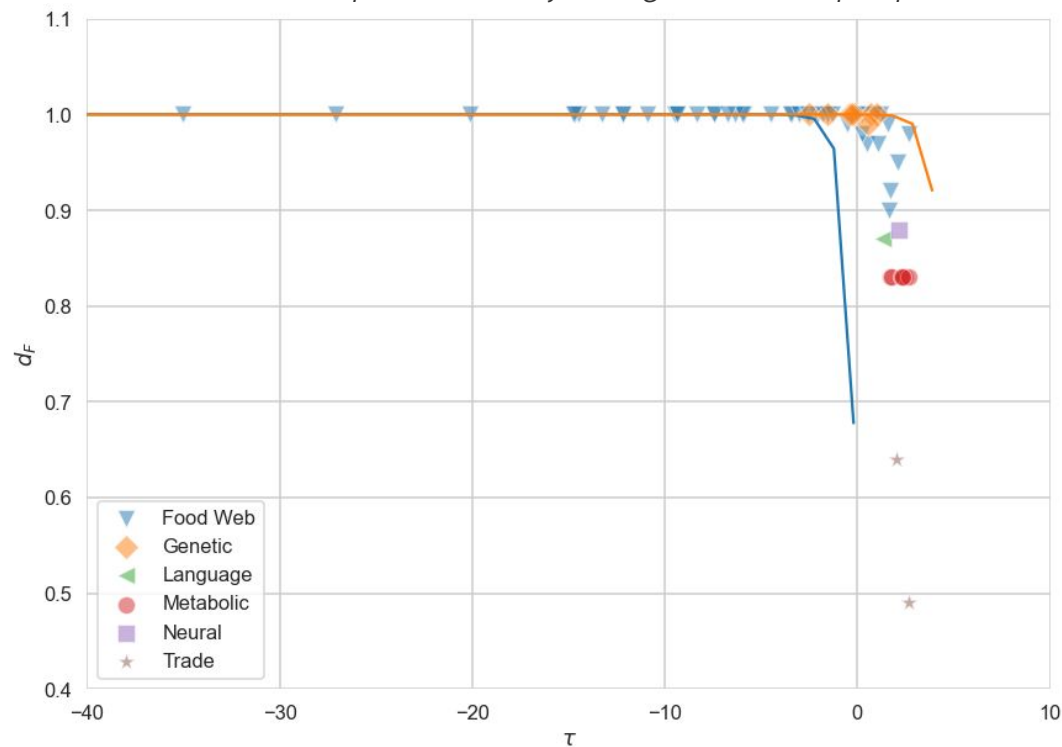
- **Strongly connected** - reach any vertex from any other vertex along directed path.
- SCC necessarily contain long cycles.
- **$\tau < 0$ - loopless regime** \rightarrow SCC is vanishingly small
- Gen-PPM displays continuous (i.e. second order) phase transition in strong connectivity with coherence

Fraction of non-basal vertices in the GC against q for the Predator-Prey model



Non-normalized: Results

Normalised deviation from normality, d_F , against the loop exponent τ

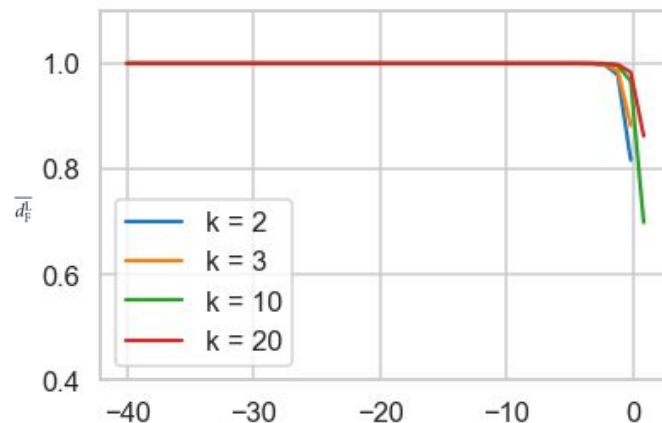


Upper and lower bounds on d_F

$$\overline{d_F^L} = \sqrt{1 - \frac{N}{L} e^{2\tau}},$$

$$\overline{d_F^U} = \sqrt{1 - \frac{1}{L} e^{2\tau}}.$$

Lower bound with different k



Real networks with small or negative τ are indeed highly non-normal

What was done:

- 1 Reproduced the results from the paper exactly.
- 2 Implemented Generalised Preferential-Prey model; Investigated the relation between T and q .
- 3 Visualised the simulated network in accordance with trophic level S .
- 4 Visualised the real complex systems network dataset in accordance with S .
- 5 Used Perceptron algorithm to re-classify networks with $\tau < 0$ and $\tau > 0$
- 6 .Compare the results obtained from Directed Configuration Ensemble with the results obtained analytically.

Conclusion

Directed networks can belong to either of two regimes which is **determine by the sign of τ** , which is a function only of trophic coherence and the in- and out-degree sequences:

Loopless



Edges are not strongly aligned with a global direction.



Trophically incoherent



Small deviations from normality



Large strongly connected components

Loopful



Edges are organized according to a global direction



Networks are highly coherent



Non-normal



Strongly connected components are vanishing



Leading eigenvalues of the datasets follow the relation

$$\overline{\lambda_1} = e^{\tau}.$$



Metabolic networks are highly incoherent and have positive in-out degree correlation in comparison wrt other networks.



Increase in q , $P(\text{acyclic})$ decreases.



Trophic coherence is parameter that is related with the non-normality and the GC are related.



Gen-PPM displays continuous (i.e. second order) phase transition in strong connectivity with coherence,



In principle, Direct configuration ensemble lead to similar results but not the same

Future Work

No. of trials-1000 - obtain mean and stdev. for each point. - q vs T and Non basal fraction vs q plots.

Only Generalized Pref-prey model implemented; Various other models could also be checked-> cascade (CM), generalized niche (GNM), niche (NM), nested hierarchy(NHM), minimum potential niche (MPNM), and PPM.

Took 1.5 hours for each point; Better parallelizing methods could be implemented since even Dask parallel computing was taking long time.

Investigate more on implementing basal ensemble with its own class of constraints.

References and Github

1. Dataset of all the networks can be found at: <https://www.samuel-johnson.org/data>
2. Johnson S, Jones NS. **Looplessness in networks is linked to trophic coherence**. Proc Natl Acad Sci U S A. 2017 May 30;114(22):5618-5623. doi: 10.1073/pnas.1613786114. Epub 2017 May 16. PMID: 28512222; PMCID: PMC5465891.
 - S I Appendix:
<https://www.pnas.org/content/pnas/suppl/2017/05/15/1613786114.DCSupplementaL/pnas.1613786114.sapp.pdf>
3. Johnson, Samuel. (2019). **Digraphs are different: Why directionality matters in complex systems**. <https://arxiv.org/abs/1908.07025v1>
4. Klaise J, Johnson S. **From neurons to epidemics: How trophic coherence affects spreading processes**. Chaos. 2016 Jun;26(6):065310. doi: 10.1063/1.4953160. PMID: 27368799.
5. Python code, datasets, csv files and figures of this project can be found at: <https://github.com/deliriarte/QLS-Project>