Data Science of Complex Systems

Semester	8
Course	ID5090
Credit Grade	0
Grade Value	0

Resources

Project Idea

Assignment 1

Assignment 2

Assignment 4

Vicsek Model

artificial “complex” system for analysis and study

$$\theta_{t+1} = < \theta_t >_{\epsilon R} + \eta_t$$

Polarization order parameter

$$\vec{P}_t = \frac{1}{N} \sum_{\forall j} \hat{\theta}_{j,t}$$

$p_t = |\vec{P}_t|$ $\quad$ $p_t = 1$ ⇒ perfectly ordered

$p_t = 0$ ⇒ fully disordered

captures how well the flock is aligned

order parameter departs from agent-view and brings in a group-view

Time series

any order parameter computed from measurements of flocking system gives rise to a time series

in most complex systems, we only have access to such order parameters and not an individual agent

A time series is “one” realization among others of the stochastic process that generates it. Our goal is to analyze and understand to uncover the actual process

Deterministic

trend & seasonality

they follow structured, explainable patterns

Trend

overall direction of time series over a long period.

increasing, decreasing or stable

Seasonality

periodic patterns that repeat at regular intervals

Stochastic

Stationarity

statistical characteristics which do not change over time

unchanging probability distribution

strictly stationary process ⇒ probability distribution does not change with different windows of time

$$f(p_1, p_2, p_3, p_4, \dots) = f(p_{\tau+1}, p_{\tau+2}, p_{\tau+3}, p_{\tau+4}, \dots)$$

weak stationarity

1. constant mean
2. constant variance
3. auto-covariance depends only on lag

$$x_t = E(p_y) = \mu \quad \forall t$$

$$\sigma_t^2 = E((p_t - \mu)^2) = \sigma^2 \quad \forall t$$

$$Cov[p_t, p_{t-k}] = E((x_t - \mu)(x_{t-\tau} - \mu)) = f(k)$$

$$\sigma_{xx}[l] = E\big((X_t - \mu)(X_{t-l} - \mu)\big)$$

Auto-correlation is given by

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}(0)}$$

For $X(t) = w(t) + aw(t-1)$

$$\begin{aligned}\sigma_{xx}(l) &= E\left[w_t w_{t-l} + a(w_{t-1}w_{t-l} + w_t w_{t-l-1}) + a^2 w_{t-1} w_{t-l-1}\right] \\ &= \sigma_{ww}(l)(1+a^2) + a(\sigma_{ww}(l+1) + \sigma_{ww}(l-1))\end{aligned}$$

$$\begin{aligned}&\sigma_{xx}(l) \\ &\qquad = (1+a^2)\sigma_w^2 \quad l = 0 \\ &\qquad = a\sigma_w^2 \qquad\qquad l = 1\end{aligned}$$

For $X(t) = w(t) + ax(t-1)$

$$\sigma_{xx}(l) = E\Big[w_t w_{t-l} + a(X_{t-1}w_{t-l} + w_t X_{t-l-1}) + a^2 X_{t-1}X_{t-l-1}\Big]$$
$$= \sigma_{wx}(l) + a\sigma_{xx}(l-1)$$

CHECK AGAIN

$$\begin{aligned} &\sigma_{xx}(l) \\ &\qquad = \frac{\sigma_e^2}{1-a^2} \quad l=0 \\ &\qquad = \frac{\sigma_e^2}{1-a^2} \quad l=1 \\ &\qquad = \frac{\sigma_e^2}{1-a^2} \quad l=2 \\ &\qquad = 0 \qquad\quad\ l>0 \end{aligned}$$

$$\begin{cases} \rho_{xx}(l) & = 1 \quad l=0 \\ & = a \quad l=1 \\ & = a^2 \quad l=2 \\ & = a^l \quad l \end{cases}$$

▼ Unnecessary

$$\left(\sum_{i=0}^{t} a^i w_{t-i}\right)\left(\sum_{i=0}^{t-l} a^i w_{t-l-i}\right)$$

$$\sigma_{wx}(0) = E(w_t * (w_t + ax_{t-1})) = \sigma_w^2$$

Time Series

$$\theta_{i,t} =< \theta_{j,t-1} >_{j \in R_I} + \eta(t)$$

$$p(t) = \frac{1}{N} \sum_k \hat{\theta}_{k,t}$$

Topology

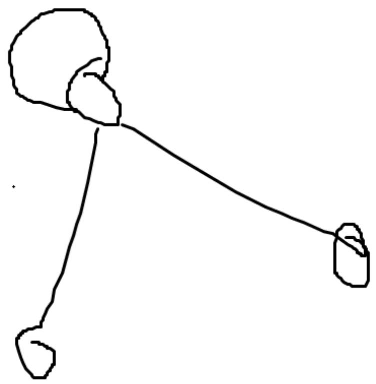
The different agents in the influece radii form a network with each other and essentially forms a graph

in the vicsek model

node	each agent
edge	when two agents are in the radius of influence of each other
weight	strength of links (same for all in vicsek)
direction	undirected

Adjacency matrix used for defining the linking between nodes and easy access

$A = A^T$



For self edge

$A_{ii} = 2$ ⇒ there is a connection from i(1) to i(2) and from i(2) to i(1)

Network

The movement affects the network and the network affects the movement

Directed ⇒ $A = A^T$

Directed ⇒ $A \neq A^T$

Density

$$
\begin{aligned}
\rho &= \frac{\#\ \text{edges}}{\text{total possible edges}} \\
&= \frac{\frac{1}{2}\sum_i \sum_j A_{ij}}{\frac{n(n-1)}{2}} \\
&= \frac{2m}{n(n-1)} \\
&\approx \frac{2m}{n^2} \qquad n >> 1
\end{aligned}
$$

where, m = number of connections

Degree of node i

$$\sum_j A_{ij} = k_i$$

Connectedness

You can either walk *(can use edge repeatedly)* or make a path *(edge is only single use)*

$$A_{ik} = \sum_J A_{ij} A_{jk} = [A^2]_{ik}$$

For a undirected matrix: $[A^2]_{ii} = k_i$

$[A^3]_{ii}$ ⇒ number of clusters *(triangle connections)* with node i

Directed Network

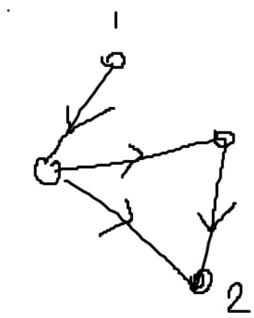
Components

we have to bring in another definition into the normal definition

weak

the nodes can only be reached in one direction

no cycles



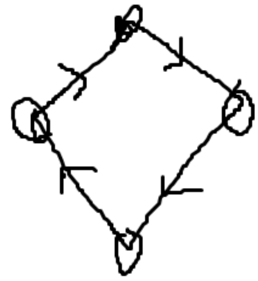
strong

the nodes can be reached in to and fro

presence of cycles

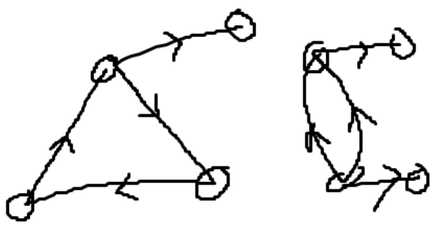
- ▼ intersection of in-component and out-component

    when you're making an in-component and out-component for a single node, it is exactly the same for any other nodes in the strong component of the former node



---

e.g.



SC: 5

    two cycles and three single nodes

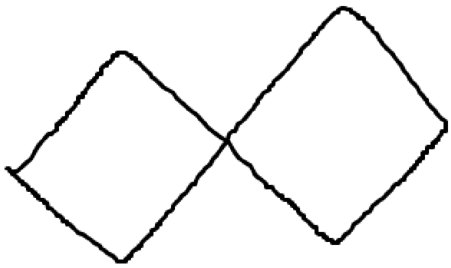
WC: 2

    left and right

Independent paths

when we talk about independent paths, there are two types

1. edge independent path
2. node independent path



node-independent path: 1

edge-independent path: 2

Random Walk on a Network

1. starts from a node i
2. at every Δt it moves to some j where $A_{ij} = 1$

$$P_i(t) \propto \sum_{j=1}^{N} \frac{A_{ij} P_j(t-1)}{k_j}$$

where $P_i(t=0) = \delta_{is}$

Arranging the above in a matrix form

$$P_t = AD^{-1} P_{t-1}$$

$$D = \begin{bmatrix} k_1 & & 0 \\ & \ddots & \\ 0 & & k_n \end{bmatrix}$$

As $t \to \infty$

$$P = AD^{-1}P$$
$$(I - AD^{-1})P = 0$$
$$(D - A)(D^{-1}P) = 0$$
$$LQ = 0$$

L ⇒ Laplacian

L is a matrix with eigen value 0 ⇒ determinant is zero for L

$$L_{ij} = \begin{cases} k_i & i = j \\ -A_{ij} & i \neq j \end{cases}$$

$Q = \overrightarrow{1}$ for $LQ = 0$

$P = DQ = \overrightarrow{k}$

$P_i = k_i$

<u>Societies</u>

<u>Equation Discovery</u>

<u>Matrix Decomposition</u>