Diffusion models on social networks

Mark C. Wilson (with Patrick Girard, Alex Raichev)
www.cs.auckland.ac.nz/˜mcw/blog/

Department of Computer Science
University of Auckland

CMSS 4th Summer Workshop, 2013-03-22

References

- M. Granovetter. Threshold Models of Collective Behavior. American Journal of Sociology 83 (6): 14201443, 1978.
- D. Kempe, J. Kleinberg, E. Tardos. Maximizing the spread of influence through a social network. Proc. KDD 2003, 137–146, 2003.
- E. Mossel, J. Neeman, O. Tamuz. Majority Dynamics and Aggregation of Information in Social Networks. arXiv:1207.0893, 2012.
- M. Jackson. Social and Economic Networks. Princeton University Press, 2008.

Motivation

- Social networks are known to have large influence on many aspects of human behaviour, including beliefs, preferences, decisions. See Jackson (2008) for basic background.

Motivation

- Social networks are known to have large influence on many aspects of human behaviour, including beliefs, preferences, decisions. See Jackson (2008) for basic background.
- Girard, Seligman, Liu developed a model of belief change in a social network.

Motivation

- Social networks are known to have large influence on many aspects of human behaviour, including beliefs, preferences, decisions. See Jackson (2008) for basic background.
- Girard, Seligman, Liu developed a model of belief change in a social network.
- This talk sets out a broader research programme.

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).
- At each time step, each agent simultaneously updates its belief, based on the beliefs of itself and its neighbours in the graph.

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).
- At each time step, each agent simultaneously updates its belief, based on the beliefs of itself and its neighbours in the graph.
- GSL model uses two rules:

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).
- At each time step, each agent simultaneously updates its belief, based on the beliefs of itself and its neighbours in the graph.
- GSL model uses two rules:
  - strong influence: if all neighbours are green, node turns green;

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).
- At each time step, each agent simultaneously updates its belief, based on the beliefs of itself and its neighbours in the graph.
- GSL model uses two rules:
  - strong influence: if all neighbours are green, node turns green;
  - weak influence: if no neighbours are red and at least one is green, red node turns yellow.

Basic features of the model

- There is an underlying connected graph $G = (V, E)$. Vertices are agents and edges represent a social relation ("friendship").
- Fix a logical proposition p. Each agent has one of three beliefs about p: true (green), false (red), undecided (yellow).
- At each time step, each agent simultaneously updates its belief, based on the beliefs of itself and its neighbours in the graph.
- GSL model uses two rules:
  - strong influence: if all neighbours are green, node turns green;
  - weak influence: if no neighbours are red and at least one is green, red node turns yellow.
- By symmetry, each rule also holds when the roles of red and green are reversed. Note that a yellow node can only change colour under strong influence.

Refining the original GSL model

- The thresholds for strong and weak influence are very high.

Refining the original GSL model

- The thresholds for strong and weak influence are very high.
- For strong influence, we can require only a strict majority of green neighbours in order to change our mind to green.

Refining the original GSL model

- The thresholds for strong and weak influence are very high.
- For strong influence, we can require only a strict majority of green neighbours in order to change our mind to green.
- For weak influence, (assuming we are red) we can become yellow if at most fraction θ of our neighbours agree with us, and at least one neighbour is green. For now, take $\theta = 0$.

The University of Auckland
NEW ZEALAND

Refining the original GSL model

- The thresholds for strong and weak influence are very high.
- For strong influence, we can require only a strict majority of green neighbours in order to change our mind to green.
- For weak influence, (assuming we are red) we can become yellow if at most fraction θ of our neighbours agree with us, and at least one neighbour is green. For now, take $\theta = 0$.
- We have a 3-colour model, with lack of symmetry, as yellow is more resistant to changing colour than the others.

Refining the original GSL model

- The thresholds for strong and weak influence are very high.
- For strong influence, we can require only a strict majority of green neighbours in order to change our mind to green.
- For weak influence, (assuming we are red) we can become yellow if at most fraction θ of our neighbours agree with us, and at least one neighbour is green. For now, take $\theta = 0$.
- We have a 3-colour model, with lack of symmetry, as yellow is more resistant to changing colour than the others.
- A related interpretation: green means “break the law”, red means “report offenders”, yellow means “stay neutral”.

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics
  - spread of disease

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics
  - spread of disease
  - adoption of new products, technologies, behaviours

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics
  - spread of disease
  - adoption of new products, technologies, behaviours
  - spread of beliefs, preferences, information

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics
  - spread of disease
  - adoption of new products, technologies, behaviours
  - spread of beliefs, preferences, information
- Abstractly, each node has a certain state (colour), and each node updates its colour based on some local rule. Updates can be simultaneous, sequential (fixed order of agents), or asynchronous (anyone can move).

Broader context

- Diffusion models on connected graphs have been widely studied. Main applications:
  - percolation in statistical physics
  - spread of disease
  - adoption of new products, technologies, behaviours
  - spread of beliefs, preferences, information
- Abstractly, each node has a certain state (colour), and each node updates its colour based on some local rule. Updates can be simultaneous, sequential (fixed order of agents), or asynchronous (anyone can move).
- There are two main methodologies: analytic results for specific network models, and agent-based simulation.

Transition rules I

- Colour changes can be deterministic or probabilistic. We focus on deterministic results in this talk. In complete generality this includes cellular automata (e.g. Conway's Game of Life) which are known to be too complex for useful study.

Transition rules I

- Colour changes can be deterministic or probabilistic. We focus on deterministic results in this talk. In complete generality this includes cellular automata (e.g. Conway's Game of Life) which are known to be too complex for useful study.
- Most common setup is 2 colours. Can use a monotone voting rule (simple game) among neighbours' colours, and switch my colour to the winning colour.

Transition rules I

- Colour changes can be deterministic or probabilistic. We focus on deterministic results in this talk. In complete generality this includes cellular automata (e.g. Conway's Game of Life) which are known to be too complex for useful study.
- Most common setup is 2 colours. Can use a monotone voting rule (simple game) among neighbours' colours, and switch my colour to the winning colour.
- In linear threshold models, each node v has a threshold $\theta_v \in [0, 1]$, and a weight b_{vw} for each neighbour w. If $\sum_w b_{vw} > \theta_v$, then v changes its colour in some way.

Transition rules I

- Colour changes can be deterministic or probabilistic. We focus on deterministic results in this talk. In complete generality this includes cellular automata (e.g. Conway's Game of Life) which are known to be too complex for useful study.
- Most common setup is 2 colours. Can use a monotone voting rule (simple game) among neighbours' colours, and switch my colour to the winning colour.
- In linear threshold models, each node v has a threshold $\theta_v \in [0,1]$, and a weight b_{vw} for each neighbour w. If $\sum_w b_{vw} > \theta_v$, then v changes its colour in some way.
- Majority dynamics falls into both classes — each node has weight 1 for all others, and $\theta_v = \deg(v)/2$.

Transition rules I

- Colour changes can be deterministic or probabilistic. We focus on deterministic results in this talk. In complete generality this includes cellular automata (e.g. Conway's Game of Life) which are known to be too complex for useful study.
- Most common setup is 2 colours. Can use a monotone voting rule (simple game) among neighbours' colours, and switch my colour to the winning colour.
- In linear threshold models, each node v has a threshold $\theta_v \in [0, 1]$, and a weight b_{vw} for each neighbour w. If $\sum_w b_{vw} > \theta_v$, then v changes its colour in some way.
- Majority dynamics falls into both classes — each node has weight 1 for all others, and $\theta_v = \deg(v)/2$.
- Higher thresholds correspond to a bias in favour of the status quo.

Transition rules II

- For m colours, I poll my neighbours, aggregate using some voting rule, and adopt the winner. Call this a local voting rule.

Transition rules II

- For m colours, I poll my neighbours, aggregate using some voting rule, and adopt the winner. Call this a local voting rule.
- Plurality is the simplest, and very little else has been studied.

Transition rules II

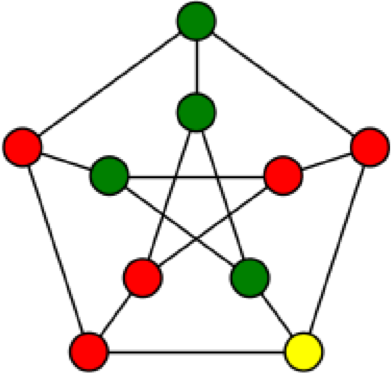
- For m colours, I poll my neighbours, aggregate using some voting rule, and adopt the winner. Call this a local voting rule.
- Plurality is the simplest, and very little else has been studied.
- GSL model is in this class, for $m = 3$. The voting rule is unnatural, and depends on the node. If I vote green, then green wins, unless everyone else votes yellow (yellow wins), or someone else votes red and no one votes green (red wins). Same with red and green reversed. If I vote yellow, then yellow wins unless everyone else votes green (green wins) or red (red wins).

The University of Auckland
NEW ZEALAND

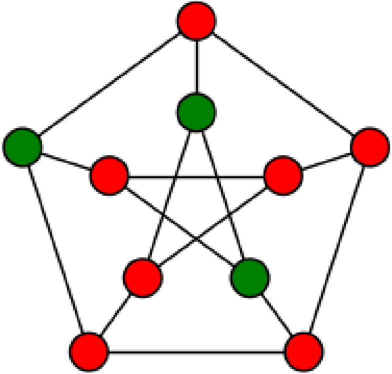
Transition rules II

- For m colours, I poll my neighbours, aggregate using some voting rule, and adopt the winner. Call this a local voting rule.
- Plurality is the simplest, and very little else has been studied.
- GSL model is in this class, for $m = 3$. The voting rule is unnatural, and depends on the node. If I vote green, then green wins, unless everyone else votes yellow (yellow wins), or someone else votes red and no one votes green (red wins). Same with red and green reversed. If I vote yellow, then yellow wins unless everyone else votes green (green wins) or red (red wins).
- The class of local voting rules that do not depend on the node, and are anonymous and neutral (that is, ordinary voting rules + polling neighbours) is worth studying.

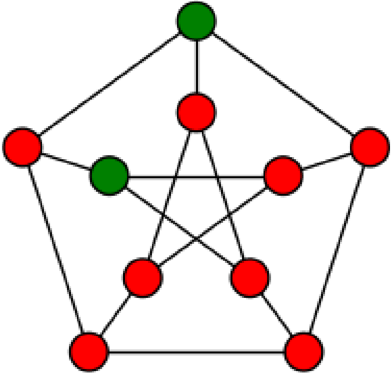
Example: Petersen, step 0



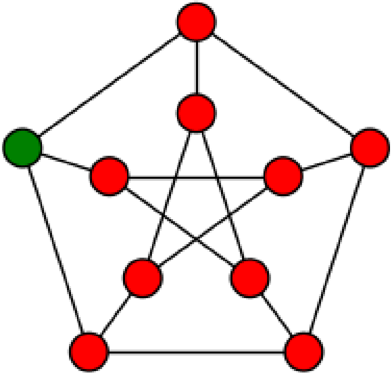
Example: Petersen, step 1



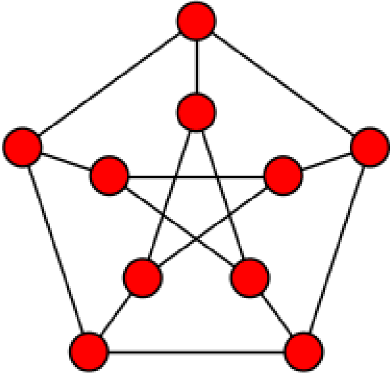
Example: Petersen, step 2



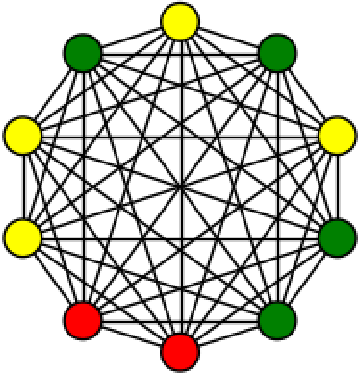
Example: Petersen, step 3



Example: Petersen, step 4



Example: K_{10}, converges immediately



Fundamental questions

- (equilibrium) Does the process converge in finite time on a given finite G? at what rate?

Fundamental questions

- (equilibrium) Does the process converge in finite time on a given finite G? at what rate?
- (unanimity) If it converges, do all nodes have the same colour?

The University of Auckland NEW ZEALAND

Fundamental questions

- (equilibrium) Does the process converge in finite time on a given finite G? at what rate?
- (unanimity) If it converges, do all nodes have the same colour?
- (wisdom of crowds) If unanimity is achieved, is it the "correct" colour? if not, does the "correct" colour win a plurality vote?

The University of Auckland NEW ZEALAND

Fundamental questions

- (equilibrium) Does the process converge in finite time on a given finite G? at what rate?
- (unanimity) If it converges, do all nodes have the same colour?
- (wisdom of crowds) If unanimity is achieved, is it the "correct" colour? if not, does the "correct" colour win a plurality vote?
- (homophily) Describe the effect on the process of assuming that nodes of same colour are more likely to be connected.

Fundamental questions

- (equilibrium) Does the process converge in finite time on a given finite G? at what rate?
- (unanimity) If it converges, do all nodes have the same colour?
- (wisdom of crowds) If unanimity is achieved, is it the "correct" colour? if not, does the "correct" colour win a plurality vote?
- (homophily) Describe the effect on the process of assuming that nodes of same colour are more likely to be connected.
- (cascades) When do arbitrary changes to some nodes propagate to a large fraction of the network?

Methodological issues

- "There are too many models, and it is hard to prove anything about them".

Methodological issues

- “There are too many models, and it is hard to prove anything about them”.
- There is no standard methodology for “network science”.

Methodological issues

- "There are too many models, and it is hard to prove anything about them".
- There is no standard methodology for "network science".
- Realistic models of social networks are hard to find. Some popular theoretical models are: Barabasi-Albert (preferential attachment), Strogatz-Watts (small worlds), Erdős-Renyi (random graph).

Methodological issues

- “There are too many models, and it is hard to prove anything about them”.
- There is no standard methodology for “network science”.
- Realistic models of social networks are hard to find. Some popular theoretical models are: Barabasi-Albert (preferential attachment), Strogatz-Watts (small worlds), Erdős-Renyi (random graph).
- Validation of models is not very advanced.

Methodological issues

- "There are too many models, and it is hard to prove anything about them".
- There is no standard methodology for "network science".
- Realistic models of social networks are hard to find. Some popular theoretical models are: Barabasi-Albert (preferential attachment), Strogatz-Watts (small worlds), Erdős-Renyi (random graph).
- Validation of models is not very advanced.
- Proving anything rigorous about dynamics, even for the above models, is technically hard.

Methodological issues

- "There are too many models, and it is hard to prove anything about them".
- There is no standard methodology for "network science".
- Realistic models of social networks are hard to find. Some popular theoretical models are: Barabasi-Albert (preferential attachment), Strogatz-Watts (small worlds), Erdős-Renyi (random graph).
- Validation of models is not very advanced.
- Proving anything rigorous about dynamics, even for the above models, is technically hard.
- Simulations may be sensitive to small changes in initial conditions.

Simple observations for GSL model

- If the strong influence threshold is sufficiently large, then any two nodes sharing an edge and having the same colour, will never change colour.

Simple observations for GSL model

- If the strong influence threshold is sufficiently large, then any two nodes sharing an edge and having the same colour, will never change colour.
- If no node is initially yellow, none will ever turn yellow. Thus the standard majority dynamics model is embedded in ours.

Simple observations for GSL model

- If the strong influence threshold is sufficiently large, then any two nodes sharing an edge and having the same colour, will never change colour.
- If no node is initially yellow, none will ever turn yellow. Thus the standard majority dynamics model is embedded in ours.
- A bipartite graph where one partition is green and the other red leads to a cycle of order 2.

Simple observations for GSL model

- If the strong influence threshold is sufficiently large, then any two nodes sharing an edge and having the same colour, will never change colour.
- If no node is initially yellow, none will ever turn yellow. Thus the standard majority dynamics model is embedded in ours.
- A bipartite graph where one partition is green and the other red leads to a cycle of order 2.
- A pendant vertex whose neighbour is not yellow will change colour immediately.

Simple observations for GSL model

- If the strong influence threshold is sufficiently large, then any two nodes sharing an edge and having the same colour, will never change colour.
- If no node is initially yellow, none will ever turn yellow. Thus the standard majority dynamics model is embedded in ours.
- A bipartite graph where one partition is green and the other red leads to a cycle of order 2.
- A pendant vertex whose neighbour is not yellow will change colour immediately.
- A star graph converges in 2 steps to unanimity if the centre is not yellow, otherwise converges immediately.

Special case: complete graph

- Homophily is irrelevant, because everyone is a neighbour of everyone else.

Special case: complete graph

- Homophily is irrelevant, because everyone is a neighbour of everyone else.
- Simultaneous majority dynamics converge in one step to unanimity if $|V|$ is odd. The crowd is wise (related to Condorcet jury theorem).

Special case: complete graph

- Homophily is irrelevant, because everyone is a neighbour of everyone else.
- Simultaneous majority dynamics converge in one step to unanimity if $|V|$ is odd. The crowd is wise (related to Condorcet jury theorem).
- With high probability, random initial colourings will converge immediately under GSL dynamics with high threshold. This happens whenever there are at least two nodes of each colour, for example.

Some related models and work

- Kempe, Kleinberg, Tardos (2003). Update rule: (generalization of) linear threshold model with $m = 2$. Aim: find optimal initial set of green nodes of given size k to maximize number of green nodes in equilibrium (influence maximization). Result: the problem is NP-complete, but a greedy algorithm gives a $(1 - 1/e)$-approximation ratio.

Some related models and work

- Kempe, Kleinberg, Tardos (2003). Update rule: (generalization of) linear threshold model with $m = 2$. Aim: find optimal initial set of green nodes of given size k to maximize number of green nodes in equilibrium (influence maximization). Result: the problem is NP-complete, but a greedy algorithm gives a $(1 - 1/e)$-approximation ratio.
- Mossel, Neeman, Tamuz (2012). Update rule: local plurality. Results: for $m = 2$, crowds are not wise in general, but they are when no orbit of the automorphism group on the graph is small.

GSL model initial explorations

- For the “right” values of parameters, three standard graph models seem to lead to consensus very often.

GSL model initial explorations

- For the “right” values of parameters, three standard graph models seem to lead to consensus very often.
- See Sage worksheet!

Plans for future work

- Concentrate on "realistic" network models, and perhaps even real networks, if we can estimate the parameters involved. Try with real data.

Plans for future work

- Concentrate on "realistic" network models, and perhaps even real networks, if we can estimate the parameters involved. Try with real data.
- Consider optimization and mechanism design problems. For example, if we can construct the network but have an upper bound on $|E|$, what topology is most likely to lead to information aggregation?

Plans for future work

- Concentrate on "realistic" network models, and perhaps even real networks, if we can estimate the parameters involved. Try with real data.
- Consider optimization and mechanism design problems. For example, if we can construct the network but have an upper bound on $|E|$, what topology is most likely to lead to information aggregation?
- Consider polling over distance-k neighbourhoods for increasing k, and quantify how phenomena depend on k.

Plans for future work

- Concentrate on "realistic" network models, and perhaps even real networks, if we can estimate the parameters involved. Try with real data.
- Consider optimization and mechanism design problems. For example, if we can construct the network but have an upper bound on $|E|$, what topology is most likely to lead to information aggregation?
- Consider polling over distance-k neighbourhoods for increasing k, and quantify how phenomena depend on k.
- Strategic behaviour - when does an agent have incentive to vote untruthfully? to break edges?

Plans for future work

- Concentrate on "realistic" network models, and perhaps even real networks, if we can estimate the parameters involved. Try with real data.
- Consider optimization and mechanism design problems. For example, if we can construct the network but have an upper bound on $|E|$, what topology is most likely to lead to information aggregation?
- Consider polling over distance-k neighbourhoods for increasing k, and quantify how phenomena depend on k.
- Strategic behaviour - when does an agent have incentive to vote untruthfully? to break edges?
- Instead of threshold dynamics, consider best reply voting, where we poll our neighbours and act as though only our 1-neighbourhood is taking part in the election. For complete graphs, some results are known.