

Article

Graph-Based Generalization of Galam Model: Convergence Time and Influential Nodes

Sining Li * and Ahad N. Zehmakan

School of Computing, Australian National University, Canberra, ACT 2601, Australia;
ahadn.zehmakan@anu.edu.au

* Correspondence: sining.li@anu.edu.au

Abstract: We study a graph-based generalization of the Galam opinion formation model. Consider a simple connected graph which represents a social network. Each node in the graph is colored either blue or white, which indicates a positive or negative opinion on a new product or a topic. In each discrete-time round, all nodes are assigned randomly to groups of different sizes, where the node(s) in each group form a clique in the underlying graph. All the nodes simultaneously update their color to the majority color in their group. If there is a tie, each node in the group chooses one of the two colors uniformly at random. Investigating the convergence time of the model, our experiments show that the convergence time is a logarithm function of the number of nodes for a complete graph and a quadratic function for a cycle graph. We also study the various strategies for selecting a set of seed nodes to maximize the final cascade of one of the two colors, motivated by viral marketing. We consider the algorithms where the seed nodes are selected based on the graph structure (nodes' centrality measures such as degree, betweenness, and closeness) and the individual's characteristics (activeness and stubbornness). We provide a comparison of such strategies by conducting experiments on different real-world and synthetic networks.

Keywords: sociophysics; Galam model; graph theory; Markov chain; social networks; convergence time; opinion formation; influential nodes; viral marketing



Citation: Li, S.; Zehmakan, A.N. Graph-Based Generalization of Galam Model: Convergence Time and Influential Nodes. *Physics* **2023**, *5*, 1094–1108. <https://doi.org/10.3390/physics5040071>

Received: 15 September 2023
Revised: 4 November 2023
Accepted: 9 November 2023
Published: 28 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Humans' decisions and opinions can be influenced by others. When forming an opinion or making a decision on a subject such as a new product or a political election, people usually seek advice from their families, friends, and other people with whom they often keep in contact with. In addition, people also consider the opinions of the figures that they value or respect; for instance, figures in specific fields or celebrities that people look up to. Therefore, people often keep updating their decisions and opinions by interacting with their connections.

Furthermore, the prevalence of online social media and social networking platforms, such as Facebook, Instagram, and WeChat, contribute to the increasing pace of opinion formation. These online platforms provide people with a convenient and fast way to communicate with friends, obtain information, and express opinions. Consequently, they significantly accelerate the information spreading and opinion formation and exchange process.

In recent decades, there has been a growing demand for a deeper understanding of how opinions form and how information spreads in social networks. A variety of opinion diffusion models, which aim to simulate the process of information spreading in a society, have thus been developed and studied within sociophysics using tools and concepts from statistical physics [1]. More recently, the topic has been enriched by contributions from theoretical computer science [2,3].

The studied models typically involve a graph G and an initial coloring of its nodes, with each node being either blue or white. The graph represents a social network, in which

each node corresponds to an individual. The edges of the graph represent relationships between individuals, such as personal contacts, friendships, or followers. The color assigned to each node, say white or blue, represents the individual's positive or negative opinion on a subject. Following the initialization, a group of nodes in the graph update their colors based on a predefined rule in each round.

The majority model is an instance of the aforementioned basic model [4,5] and it has become popular within the community of social networks. In this model, all nodes in the graph G simultaneously update their color to the most frequent color among their neighbors. If there is a tie, a node does not change its current color. It should be noted that the number of neighbors of each node can be quite large, which makes the model somewhat unrealistic. For example, it is almost impossible for a person who has 50 friends to discuss a topic with all the friends at the same time.

Another established model is the Galam model [6], which describes the dynamics of the spreading of the minority opinion in a social debate. In this model, all individuals (nodes) are randomly assigned to groups of different sizes, based on the geometry of social life of people, with the discussing groups composed of small sizes from 1 person up to 5 or 6 people. Only the people in the same group can discuss the topic and update their opinion based on the local majority opinion in the group. Therefore, the Galam model is more realistic. In the general majority-based model, the nodes with a higher degree can influence more nodes in each round. But, in the Galam model, the node can only influence other nodes in the same group. Moreover, at a tie, the Galam model introduces the possibility of a tie breaking determined using the unconscious prejudices of the agents [7].

In the classic Galam model, each individual has the same probability of grouping with any other individual. However, it is unusual that two people get together and discuss on some topics if they do not have any relationships. To add another layer of realism to the model, we leverage the underlying graph structure of a social network to introduce a generalization of the Galam model in the present paper. In the Graph-based Galam model (GGM), only the nodes that form a clique in the graph can be assigned to a group.

In the present paper, we focus on two problems: (1) How long does it take for the GGM to converge? This is arguably the most well-studied question for different dynamic systems. In our model, the answer highly depends on the underlying graph structure. (2) What is the subset of nodes which maximizes the final expected number of blue nodes, if initially blue, for a given fixed budget? This is also another popular question investigated for various models, motivated from viral marketing. For example, when a company plans to launch a new product on to the market, it is often intended to persuade a specific group of people into using the product to maximize the final cascade of the adoptions of the product. Another example is when a political party aims to convince a subset of individuals to adopt a positive opinion about a topic with the aim of a large further adoption of the positive opinion. The main problem here is to find a subset of nodes (individuals) of a given size whose adoption of the desired product/opinion results in its maximum adoption at the end.

We study the convergence time of GGM on different synthetic and real-world graph data. Experiments conducted on cycle and complete graphs show that the convergence times of GGM on cycle and complete graphs are in $\mathcal{O}(n^2)$ and $\mathcal{O}(\log n)$, respectively, where n denotes the number of nodes. We also investigate how adding new edges can influence the convergence time of the process and observe that it essentially depends on the manner that the edges are added. Particularly, if the added edges increase the expansion (connectivity) of the graph, the convergence time usually decreases.

Furthermore, the experiments on complete graphs and real-world social networks indicate that the initial blue nodes must be more than half to ensure that the blue nodes "dominate" at the end, if one selects the initial blue nodes randomly. If the initial blue nodes are selected based on some smarter strategies such as centrality-based measures (such as degree, betweenness, and closeness), then a much smaller fraction of initial blue nodes is required for the blue color to be the winning color at the end.

Inspired by some prior work [8,9], we also study the setup where the nodes have different levels of activeness and stubbornness; that is, some nodes are more probable to participate in the gatherings and some nodes are less inclined to follow the majority opinion. We then introduce some strategies for choosing initial seed nodes, which take these parameters into account, in addition to the graph structural properties.

The paper is organized as follows. We first give some basic definitions and an overview of some prior work in the rest of this Section. Our findings, which address questions (1) and (2), are presented in Sections 2 and 3, respectively.

1.1. Preliminaries

1.1.1. Graph Definitions

- Graph

Let $G = (V, E)$ be a simple connected undirected graph. Let $n := |V|$ and $m := |E|$ represent the number of nodes and the number of edges, respectively. For a given node, $v \in V$, $N(v) := \{u \in V : \{u, v\} \in E\}$ represents the neighborhood of v . Furthermore, let $d(v) := |N(v)|$ represent the degree of node v .

- Expander Graph

An expander graph is expected a graph with good connectivity. There are three parameters often being used to measure the expansion of a graph: vertex expansion, edge expansion, and spectral expansion. A graph will be called here an expander graph if it has strong expansion properties.

(1) Vertex Expansion. Let $S \subseteq V$ be a subset of nodes, and $\bar{S} := V \setminus S$. Let $\partial S_{\text{out}} := \{v \in \bar{S} | \exists u \in N(v), \text{ such that } u \in S\}$, which is the outer boundary of set S . The vertex expansion is defined as $h_{\text{out}}(G) := \min_{0 < |S| < \frac{n}{2}} |\partial S_{\text{out}}| / |S|$. For nodes in any “small” subset S (with a size less than $n/2$) of V , the greater the $h_{\text{out}}(G)$ is, the larger the number of their neighbors outside S is, and the better connected the graph G is. Therefore, a graph G with a greater $h_{\text{out}}(G)$ has stronger expansion properties

(2) Edge Expansion. Let $S \subseteq V$ be a subset of nodes, and $\bar{S} := V \setminus S$. Let $\partial S := \{\{u, v\} \in E | u \in S, v \in \bar{S}\}$, which is the boundary of set S . The edge expansion is defined as $h(G) := \min_{0 < |S| < \frac{n}{2}} |\partial S| / |S|$. For the number of edges from the nodes inside any “small” subset (with a size less than $n/2$) of V to the nodes outside this subset, the greater the $h(G)$ is, the larger the number of edges is, and the better connected the graph G is. Therefore, a graph G with a greater $h(G)$ has stronger expansion properties

(3) Spectral Expansion. Let A be the adjacency matrix of the graph G . Since A is symmetric and real, it has n real-valued eigenvalues. Let $\lambda(G)$ represent the second-largest absolute eigenvalue of A . A graph G with a smaller value of $\lambda(G)$ has stronger expansion properties.

1.1.2. Model

- Majority Model and Random Majority Model

Consider a graph G that represents a social network. Each node in the graph is either blue or white at the initial state, which represents a person holding a positive or negative opinion about a product or a topic. In each round, all the nodes simultaneously update their color to the most frequent color among their neighbors. If there is a tie, the node keeps its color. This is known as the Majority Model. The Random Majority Model is the same as the Majority Model, except for the tie-breaking rule. In the Random Majority Model, a node chooses blue or white with an equal probability of 0.5 in case of a tie.

- Galam Model

Consider a population of n individuals (nodes) who can hold a positive or negative opinion on a subject (i.e., are blue or white). Furthermore, consider a set of rooms of various sizes, such that the summation of the size of the rooms is equal to n . In each

round, all individuals are randomly assigned to these rooms. Then, all individuals simultaneously update their opinion to the most frequent opinion in the room. If there is a tie, then a tie-breaking rule is applied to handle the situation. In the original description of the model, individuals choose negative in case of a tie [6]. However, other variants are considered, for example, random tie-breaking rules [7].

- Graph-based Galam Model

In the present paper, we introduce a graph-based generalization of the Galam model, GGM. A graph $G = (V, E)$ is used to represent the society to be considered. The nodes and edges in the graph represents the individuals and the relationship between them, respectively. We define a coloring function $C: V \rightarrow \{w, b\}$, where w and b represent white and blue, respectively, which correspond to negative and positive.

There are two steps in each round of this model: (1) randomly assign all nodes into groups with different sizes; and (2) update the color of each node following the local majority-based rule in the group.

- (1) Group Allocation. All the cliques of size 1 to R in the graph are collected and stored in a list. In our setup, we use $R = 3$, but the model is well defined for larger values of R . Each time, one clique in the list is randomly picked with an equal likelihood. For each node in the clique that is picked, the remaining cliques that contain the node is removed from the clique list. This continues until the nodes are partitioned into cliques of size 1, 2, and 3 (the clique list is empty).
- (2) Color Updating. All the nodes simultaneously update their color to the most frequent color in their group. If there is a tie, a binary random number (0 or 1) is generated with equal probability. The nodes becomes blue if the random number is 1 and white otherwise.

An example is given in Figure 1. The graph on the left-hand side is the initial state. The list of all cliques of this graph is $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \{1, 5\}, \{2, 5\}, \{3, 4\}, \{4, 5\}, \{4, 6\}, \{5, 6\}, \{1, 2, 5\}, \text{ and } \{4, 5, 6\}\}$. One possible outcome of group allocation is $\{\{1, 2\}, \{3\}, \text{ and } \{4, 5, 6\}\}$. In this case, node 1 and node 2 have the same color and keep their color. Node 3 is in a group of itself and it also keeps the color. Nodes 4, 5, and 6 are in the same group, and the major color in this group is blue. Hence, node 4 updates its color from white to blue. The state of the graph after this round is illustrated in the graph in Figure 1, right.

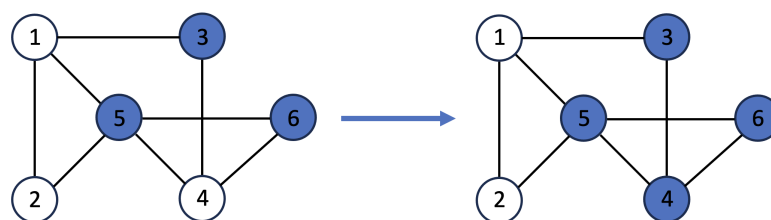


Figure 1. An example of one round of the Graph-based Galam model: the initial state of the graph (**left**) and the updated graph after one round (**right**).

1.2. Prior Studies

- Models

Numerous models that simulate the spreading of information and the formation of opinions have been introduced and studied, such as the Independent Cascade (IC) model [10], the Linear Threshold (LT) model [10,11], and majority-based models [12–15]. The Galam model is one of the most studied models in the area of sociophysics. The model was originally designed to explain how an initial minority can finally win the debate [6]. This model has become an established model in sociophysics and various variants of this model have been investigated. Some of these variants involve considering three different opinions [16], changing the biased tie-breaking rule to a random one [7], adding inflexible individuals [8], and introducing the level

of activeness [9]. The model studied in this paper falls under the umbrella of this line of models.

- **Convergence time**
Svatopluk Poljak and Daniel Turzík proved that the convergence time of the majority model is upper bounded by a linear function of the number of edges; that is, $\mathcal{O}(m)$ [17]. Furthermore, some stronger bounds have been proven for some special types of graphs; see [18]. Studies on convergence property have also been conducted for other majority-based models [19,20]. For the random majority model on a cycle graph, the convergence time is proven to be in $\mathcal{O}(n^2)$ [18]. For the classic Galam model, Bernd Gärtner and one of the authors of this paper proved that the convergence time is in $\mathcal{O}(\log \log n)$ when all groups in the model have a size greater or equal to 3 [21].
- **Influential nodes**
The research about viral marketing in social networks has become popular in recent decades. A small set of seed individuals who hold a positive (blue) opinion on some subject needs to be selected to maximize the final number of individuals holding a positive opinion in the social network. Such a problem about how to select seed individuals is typically known as influence maximization (IM). David Kempe, Jon Kleinberg and Éva Tardos provided a foundation for this problem in their seminal paper in 2003 [10]. After that, there has been extensive study on the IM problem [11,22–24].
The above IM problem can be modeled as a discrete optimization problem and it is usually proven to be computationally hard to solve for various models; more precisely, it is known to be non-deterministic polynomial-time hard (NP-hard); see [10,24]. Therefore, several approximation, randomized, and heuristic algorithms, such as centrality-based algorithms [25,26] (which choose nodes with the highest degree, betweenness, closeness, or pagerank) and greedy approaches [27–29], have been proposed. In recent years, machine learning techniques have become popular, which leads to the development of some machine learning-based algorithms for the IM problem [30–32].

1.3. Experimental Setup

All graph data for the experiments are from Stanford Network Analysis Project (SNAP) [33]. Experiments were conducted on Facebook (FB) and Twitch Spain (T-ES)’s social networks. Some basic properties of these graphs are listed in Table 1. Furthermore, several experiments focus on cycle graphs and complete graphs to examine the relationship between the number of nodes and the convergence time.

Table 1. Basic properties of the two examined social networks. Here, n , m and “AvgDegree” denote the number of nodes, the number of edges and the average degree, respectively.

Social Network Name	n	m	Avg. Degree
Facebook	4039	88,234	43.69
Twitch Spain	4638	59,382	25.55

Since our model is a random process, all experiments were executed 20 times, unless otherwise specified, and the average output was considered. All experiments were conducted on an Apple M1 Pro chip, with 32 GB RAM and MacOS Ventura (Apple Inc., Cupertino, CA, USA).

2. Convergence Time

Some experiments were conducted to study the convergence time of the GGM in cycle graphs, complete graphs, and some other types of special graphs in this Section. Before discussing the outcomes of these experiments, we show that the GGM always converges on a connected graph.

Convergence to an Absorbing State. For a connected graph G with n nodes, the model on such a graph can be regarded as a Markov chain. Since there are 2^n possible colorings in the model, this Markov chain has 2^n states. If the probability of transition from state s to s' is non-zero, then there is a directed edge from s to s' . A state is called an absorbing state if there is no edge going out, and a Markov chain is called an absorbing Markov chain if every state in the Markov chain can reach an absorbing state.

In GGM, the two states where all nodes have the same color (blue or white) are two absorbing states. Since each node in the graph keeps its color in such two states, there is no outgoing edge from the two states. Except these two states, any other state has at least two edges going out from it and there is a path from the state to the absorbing states. Consider a state that there are i blue nodes ($i \neq 0$ and $i \neq n$, i.e., there are not n blue nodes or n white nodes). Since the graph is connected, there must be at least two adjacent nodes with different colors. In the coming round, there is a non-zero probability that these two nodes are assigned to the same group and all other nodes are assigned to other groups of size one. Then, these two nodes become both blue or white and all other nodes keep their color at the end of this round. Therefore, the state that has i blue nodes must have at least two outgoing edges: one edge goes to a state that has $(i + 1)$ blue nodes and the other one goes to a state that has $(i - 1)$ blue nodes. As the process keeps going, the state of i blue nodes finally reach the state that all the nodes are blue (or white). Hence, the GGM on a connected graph is an absorbing Markov chain with two absorbing states where all nodes are blue or white. Furthermore, the process eventually converges to one of the absorbing states.

2.1. Cycle Graph

As mentioned in Section 1.2, prior studies have shown that the convergence time of the random majority model on a cycle graph with n nodes is in $\mathcal{O}(n^2)$. Since one can think of the GGM somewhat as a room-based random majority-based model, let us speculate that the convergence time of the model in a cycle graph with n nodes is also of order $\mathcal{O}(n^2)$ for GGM.

The experiments were conducted on 10 cycle graphs with the number of nodes selected from 100 to 1000 every 100. The number of initial blue nodes and how to set these blue nodes should be determined to maximize the convergence time. One can expect that it would take more time for the GGM to converge when the number of blue and white nodes are equal. If blue nodes are more (or less) than white nodes, the process tends to converge to a state of all nodes being blue (or white) more quickly. Therefore, the initial number of blue nodes were set to be half of the total nodes. In addition, the blue nodes were selected in two ways. The first way is that all the blue nodes are next to each other to form a path. The other way is that all blue nodes and white nodes are alternating in the cycle. An example of these two types of initial state can be seen in Figure 2. For the graph in Figure 2a, only four nodes have a non-zero probability to change their color in the coming round. However, all the nodes in the graph shown in Figure 2b have a non-zero probability to change their color. All the other configurations of initial blue nodes are expected to place between these two special cases, with regard to convergence time.

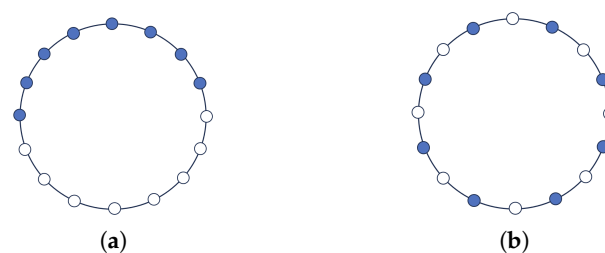


Figure 2. An example of the initial state of a cycle: (a) all blue nodes are next to each other; (b) white and blue nodes are alternating in the cycle.

Figure 3 shows the outcome of the experiments. Cycle graph (1) and cycle graph (2) correspond to the two ways of initializing. The experiment result indicates that the convergence time of the GGM on cycle graphs is of order $\mathcal{O}(n^2)$. When half of the nodes are blue, the two different ways of setting these blue nodes have little impact on the convergence time.

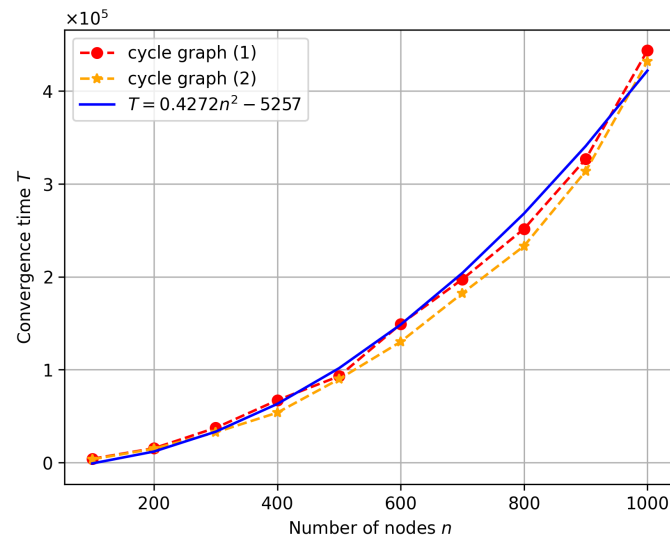


Figure 3. Convergence time in cycle graph.

2.2. Complete Graph

Prior study [21] has shown that the convergence time of the original Galam model with all groups of a size greater or equal to 3 is of order $\mathcal{O}(\log \log n)$. In the original Galam model, any two individuals in the social network can be assigned in the same group, which means that there exists a relationship between them. If a graph is used to represent the social network, there must be an edge between any two nodes in this graph. Then, the original Galam model can be regarded as a complete graph. Therefore, one might speculate that the convergence time of the GGM on complete graphs with n nodes is of order $\mathcal{O}(\log \log n)$.

The experiment was conducted on 10 complete graphs with the number of nodes selected in the set $\{50, 100, 200, 500, 1000, 2000, 3000, 4000, 5000, \text{ and } 10,000\}$. At the initial state, half of the nodes were blue, since as expected, this would result in the largest convergence time. Since each node in a complete graph is connected to all other nodes and there is no structural difference between the nodes, the blue nodes were just randomly selected.

The outcome of the experiment is depicted in Figure 4a. It should be noted that the convergence time of the GGM on complete graphs is of order $\mathcal{O}(\log n)$, but not $\mathcal{O}(\log \log n)$. The main reason for the difference is more expected to be the tie-breaking rule and the number of groups of even size. In the original Galam model, all the people in the group hold a negative (white) opinion if there is a tie. It is a biased tie-breaking rule, which can lead to fast convergence if there are more groups with a size of an even number. In addition, the convergence time $\mathcal{O}(\log \log n)$ has been proved when all groups have a size of equal to or greater than 3. However, the group size in the GGM in the present paper is less than or equal to 3, which might slow down the diffusion of opinions.

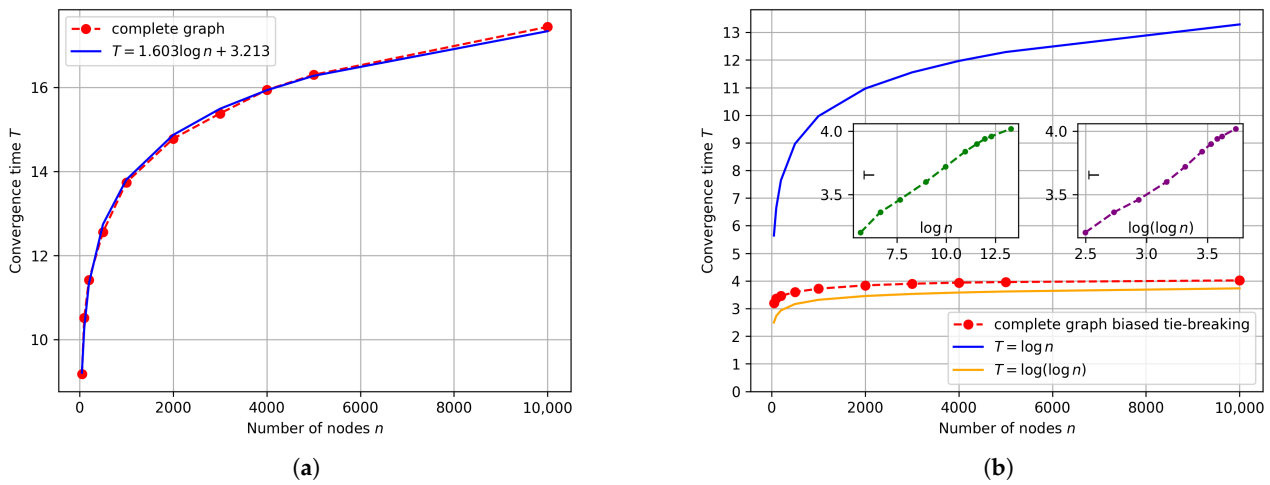


Figure 4. Convergence time in complete graph with (a) random tie-breaking rule and (b) biased tie-breaking rule.

We conducted another experiment on those 10 complete graphs with the biased tie-breaking rule. The maximum size of a group was increased from 3 to 4. Figure 4b shows the result of this experiment. There are two insets in Figure 4b, which illustrate the plots of the convergence time, T , against $\log n$ and $\log(\log n)$. The insets show that the convergence time is a linear function of both the logarithm and double logarithm of n . The reason is that $\log(\log n)$ is a linear function of $\log n$ when n is not big enough. Thus, experiments should be executed for n that is large enough to determine the order of the convergence time. However, we cannot run the experiments for so large n due to the limitation of computing power. In such a case, we plot the curves of the convergence time T against the number of nodes n , $T = \log(\log n)$ and $T = \log n$ in Figure 4b to make a direct comparison. It can be observed that the curve of T against n is much closer to the curve of $T = \log(\log n)$. It indicates that the convergence time of the GGM on a complete graph looks to be rather of order $\mathcal{O}(\log \log n)$ if the biased tie-breaking rule is followed and the maximum group size is increased and is an even number.

2.3. Some Other Special Graphs

The outcomes of the experiments on the cycle and complete graphs from the previous sections show that the GGM converges much faster on a complete graph than on a cycle graph if they have the same n . For an n -node graph, the complete graph has $\binom{n}{2} = n(n - 1)/2$ edges, but a cycle graph only has n edges. One may expect that adding more edges to a graph can increase the convergence rate.

An experiment was executed on two-cycle graphs (Figure 5a) and random cycle graphs (Figure 5b). A two-cycle graph can be generated by taking a cycle graph and adding an edge between every two nodes with a distance of 2. To build a random cycle graph, two randomly selected extra edges should be added to each node in a cycle graph.

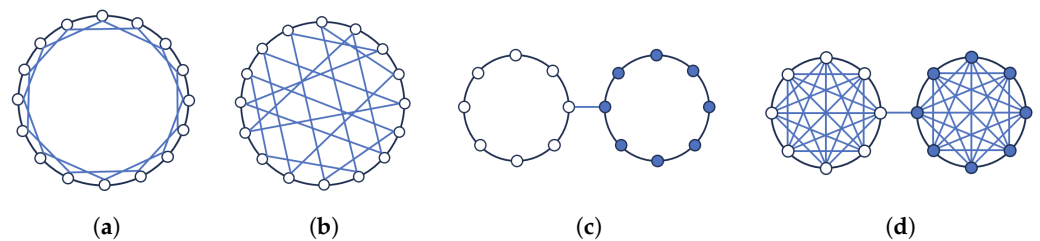


Figure 5. (a) Two-cycle, (b) random cycle, (c) two-cycle-connected and (d) two-complete-connected graphs.

Figure 6a,b depicts the convergence time of the GGM on two-cycle graphs and random cycle graphs, respectively. It can be observed that the process converges faster on two-cycle graphs than on a cycle graph, but the convergence time is still of order $\mathcal{O}(n^2)$. However, the convergence time is of order $\mathcal{O}(n)$ on the random cycle graph. With the same n and m (number of edges), a two-cycle graph converges much slower than a random cycle graph. Therefore, the increase in the convergence speed is not merely the result of adding extra edges, but rather how to add these edges.

It should also be noted that adding extra edges improperly can even slow down the convergence of the process. Figure 5c,d illustrates the two-cycle-connected (TCC) graph and the two-complete-connected (TKC) graph, respectively. These types of graphs can be built by connecting two cycle or complete graphs of the same size with one edge. It is obvious that the number of edges in a TKC graph is greater than that in a TCC graph if they have the same n . We run experiments on these two types of graphs. The outcome shows that the convergence time of the GGM on a TCC graph is $\mathcal{O}(n^2)$ (Figure 6c). However, the TKC graph has a convergence time of $\mathcal{O}(2^{n/2})$ (Figure 6d) even though it has more edges than a TCC graph with the same n . It takes much longer for a TKC graph to converge (i.e., more than 200,000 rounds for only 30 nodes).

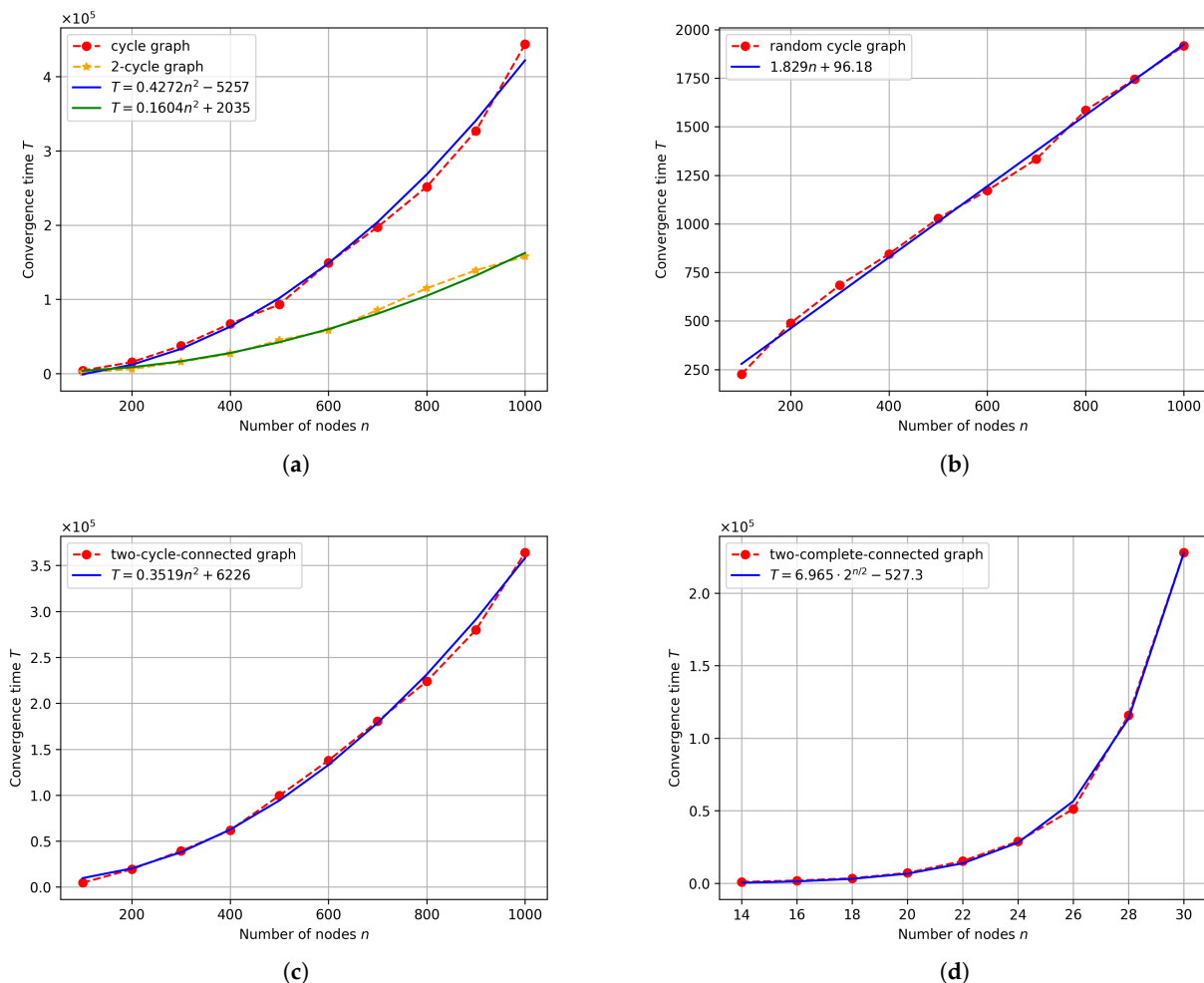


Figure 6. Convergence time in (a) two-cycle, (b) random cycle, (c) two-cycle-connected, and (d) two-complete-connected graphs.

Considering the expansion property of TCC and TKC graphs, both of them have an edge expansion of $2/n$ and a node expansion of $2/n$. It means these two types of graphs have a poor connectivity (removing only one edge or one node can make them become disconnected). The spectral expansion of a TCC graph with 50 nodes is only 2.236 and

it is much smaller than the spectral expansion of a 50-node TKC graph, which is 23.962. Therefore, the TCC graph has better expansion properties than the TKC graph.

Furthermore, in the case of a two-cycle and random cycle graph, while they have the same number of nodes and edges, the random cycle graph has much stronger expansion properties since the randomly added edges significantly improve the connectivity of the graph. Strong expansion and connectivity properties can increase the flow of information in the network. Consequently, the process is much faster on a random cycle graph than on a two-cycle graph, as observed in our experiments.

From the results of these experiments, it can be observed that the expansion property has an impact on the convergence time. For instance, a complete graph or random cycle graph is a good expander and the process in such a graph converges fast. One can expect that the convergence time of the GGM on a graph with strong expansion properties to be less.

3. Influential Nodes

As discussed in Section 2, the original Galam model can be modeled by a complete graph. Since there is no difference between each node in a complete graph, only the number of initial blue nodes influences the total number of blue nodes at the end. Thus, the initial blue nodes are randomly selected. Experiments were conducted for the GGM on a 1000-node complete graph with different initial ratios of blue nodes selected in the set $\{0.1, 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, \text{ and } 0.9\}$. Experiments were executed 1000 times for each initial ratio of blue nodes. When the process converges to an absorbing state, the average final ratio of blue nodes is checked. Figure 7 shows that the GGM converges to the absorbing state that all nodes are blue (white) eventually if the initial ratio of blue nodes is just over (under) half. This looks to be true in any graph, which possesses some level symmetry (such as vertex-transitive graphs).

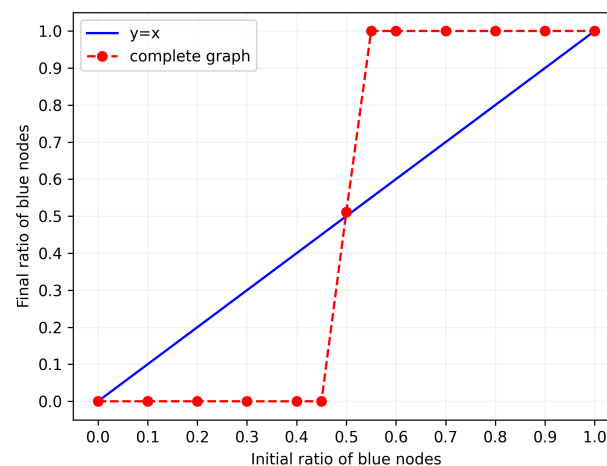


Figure 7. Randomly select initial blue nodes on a 1000-node complete graph.

However, most real-world social networks cannot be modeled as a complete graph. For instance, some individuals may have more connections to a social network than others. Such individuals are modeled as nodes with a higher degree. Selecting these nodes as initial blue nodes may lead to different results compared to selecting nodes with a lower degree. Therefore, a strategy should be developed to smartly choose the initial blue nodes rather than just randomly choosing them. One of the most often used approach is to use different centrality based measures.

3.1. Centrality-Based Influential Nodes

The centrality of a node can be used to identify the “importance” of the node [34]. There are many definitions of centrality, and the degree, betweenness, and closeness centrality are considered in the present paper.

Degree. As defined in Section 1.1.1, it is the number of neighbors of a node:

$$d(v) := |N(v)|.$$

Betweenness. If a node lies on as many shortest paths between other nodes in a graph, it is more probable that this node is placed in the center of the graph. Based on this idea, the betweenness centrality is defined as

$$b(v) := \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}.$$

where σ_{st} is the total number of shortest paths from node s to node t , and $\sigma_{st}(v)$ is the number of those paths passing through the node v .

Closeness. If the total distance between a node and all other nodes is short, then the node is close to all other nodes. The closer a node to all other nodes, the more central the node is. The closeness centrality comes from this idea and it is defined as

$$c(v) := \frac{n-1}{\sum_u d(u,v)}.$$

where $d(u,v)$ is the distance (shortest path) between node u and node v .

To compare the difference between randomly selecting initial blue nodes and choosing initial blue nodes based on centrality, we created three lists for the graph on which our model ran. Each list stores all the nodes of the graph. Then, these lists were sorted in descending order of the degree, betweenness, and closeness centrality of the nodes, respectively. When setting the initial blue nodes, we choose the first $\lceil \alpha n \rceil$ (α is the ratio initial of blue nodes) nodes in the list as blue nodes.

We ran our model on the T-ES and FB social networks. Four experiments were designed on each social network. One experiment is for randomly selecting the initial blue nodes and the other three are for selecting the initial blue nodes according to the three types of centrality. It should be mentioned that there is often a time limit for a product to take over a market in the real world. We assumed that each person takes part in three discussions every day on average, and the time limit is 100 days. Therefore, we ran the GGM 300 rounds in each experiment instead of waiting for it to converge. Note that running the process until the final convergence can also be quite expensive computationally for these networks.

Figure 8 shows the outcomes of these experiments. When the initial blue nodes are randomly selected, the outcomes of the experiments on these two real-world social networks are quite similar to that on the complete graph, where 0.5 is a transition point. The final ratio is more than 0.9 (less than 0.1) and tends to be 1 (0) when the initial ratio is greater (less) than 0.5. On the other hand, if the initial blue nodes are selected according to the centrality-based measures, then even a smaller fraction of initial blue nodes can result in a final blue ratio that is greater than 0.5.

For the T-ES social network, there is not much difference between choosing the initial blue nodes based on degree, betweenness, and closeness centrality. When the initial ratio is equal to or greater than 0.4, the final ratio is close to 1. For the FB social network, the difference between degree-based and betweenness-based strategies is not big. However, the initial and final ratios of blue nodes are almost the same when the initial blue nodes are selected according to their closeness. In fact, the number of blue nodes just oscillated around the initial number during the experiment. This may be attributed to the structure of FB’s social network.

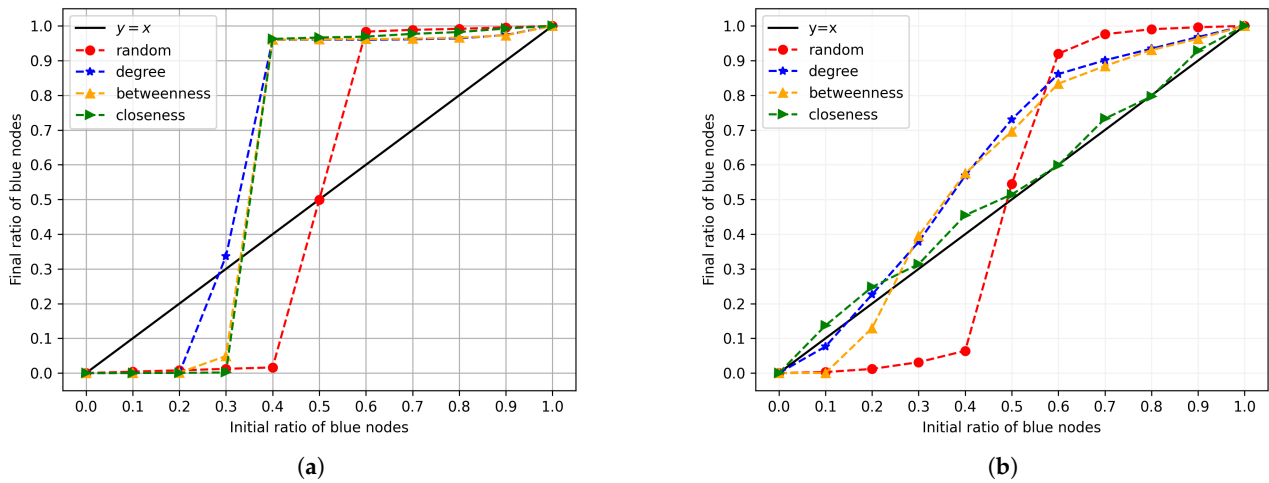


Figure 8. Centrality-based influential nodes selection on (a) Twitch Spain’s and (b) Facebook’s social networks.

3.2. Personality-Based Influential Nodes

In addition to considering how “central” individuals in a social network are, one also needs to take into account their personality when picking up the initial blue nodes. For instance, some people are more active and they spend more time on social communication. These people are more expectedly to influence their connections than those who communicate little with others. In addition, some individuals are inflexible and they firmly stick to their vision. To maximize the final ratio of blue nodes, it might be a reasonable way to choose the initial blue nodes that are more active and inflexible.

In Refs. [8,9], the “inflexible agent” and “activeness” have been introduced. In the present paper, we re-introduce these two parameters, called “activeness” and “stubbornness”, for each node in a graph and define them in a slightly different way (which we believe is more self-explaining). Each node is assigned a value between 0 and 1 for its activeness and a value between 0 and 1 for its stubbornness. In our experiments, two random numbers were generated from the uniform distribution (0, 1) independently and they were set as activeness and stubbornness, respectively. For a node, the greater the activeness (stubbornness) is, the more active (inflexible) is the node. More precisely, a node v with activeness $act(v)$ decides to participate in the opinion exchange process with probability $act(v)$. Furthermore, for stubbornness $stub(v)$, it decides not to follow the majority updating rule and keep its opinion unchanged.

The GGM needs to be modified after these two parameters are introduced. Before each round of the model, a random number following uniform distribution (0, 1) to be generated for each node in the graph. If the random number is greater than the activeness of a node, then the node to be removed from the graph in this round. When each node is updating their colors, a random number following uniform distribution (0, 1) is generated for each group. If the stubbornness of a node in the group is greater than the random number, this node keeps its color, even though its color is not the majority in the group.

Before experiments on the modified model were executed, three lists of all nodes were created. These three lists are sorted in descending order of the degree, activeness, and stubbornness, respectively. The first $[an]$ was selected as the initial blue nodes.

The results are illustrated in Figure 9. It can be observed that selecting the initial blue nodes according to the stubbornness of nodes always maximizes the final ratio of blue nodes in T-ES social networks. In FB social network, the stubbornness-based strategy maximizes the final ratio of blue nodes when the initial ratio is greater than or equal to 0.4.

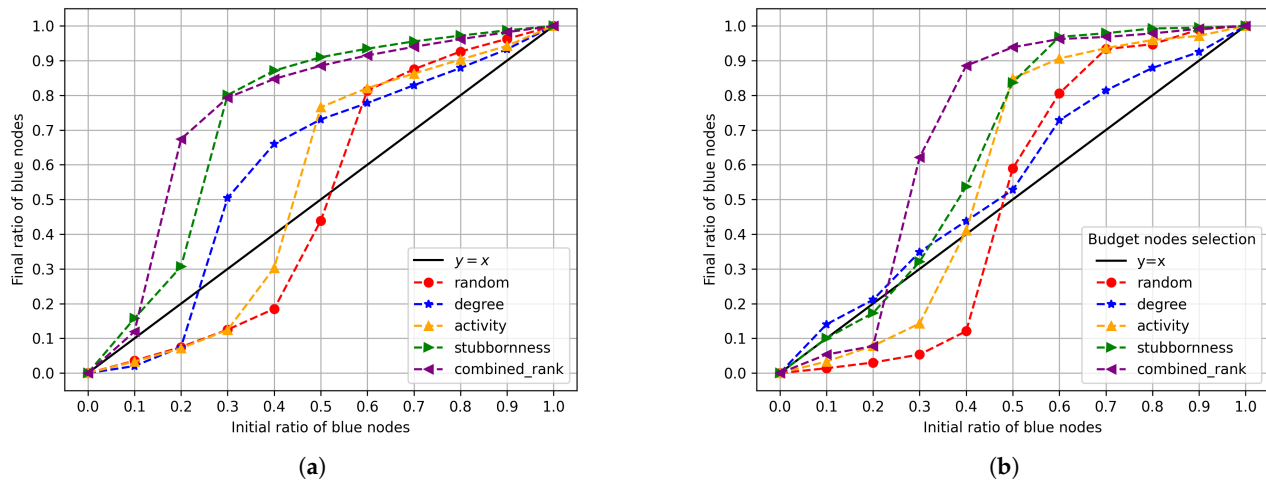


Figure 9. Personality-based influential node selection on (a) Twitch Spain’s and (b) Facebook’s social networks.

The results indicate that the stubbornness is the dominant parameter in this model. But, what about selecting the initial nodes based on a rank that combines the centrality and personality of a node? A new parameter called combined rank is introduced, considering $d(v)$, $act(v)$, and $stub(v)$. As stubbornness has more of an impact on the final ratio of blue nodes than the other two factors, more weight are added to the $stub(b)$ in the combined rank. Thus, the combined rank $r(v)$ is defined as

$$r(v) = \sqrt{d(v)} \cdot \sqrt{act(v)} \cdot (stub(v))^3.$$

We ran another experiment in which the initial blue nodes were picked up based on the combined rank. The outcomes are the purple line in Figure 9a,b. It can be observed that only 20% of the initial blue nodes will result in about 70% of blue nodes at the end in the T-ES social network. For FB social network, 30% of the initial blue nodes result in about 60% of blue nodes.

4. Conclusions

In the current paper, we generalized the classic Galam opinion formation model by using a graph structure to represent a social network. Conducting several experiments on some special types of graphs, we showed that the convergence time of the GGM on the cycle and complete graphs are in $\mathcal{O}(n^2)$ and $\mathcal{O}(\log n)$, respectively, and it is probable that the expansion of a graph can influence the convergence time of this model. The model on a graph with strong expansion properties (e.g., any “small” subset of vertices of the graph has a great amount of connections with the complement of the subset, or the second-largest eigenvalue of the adjacency matrix of the graph is small) is expected to converge faster. Finding an explicit relationship between the expansion of a graph and the convergence time of GGM could be an interesting problem to tackle in the future.

Furthermore, experiments on real-world social networks indicate that selecting the initial blue nodes based on their centrality properties and personalities can lead to more final blue nodes compared with choosing the initial blue nodes randomly. It should be noted that the number of blue nodes just oscillated around the initial number when we select the initial blue nodes according to the closeness centrality on FB’s social network. What types of graph structure can lead to such an outcome? Why does the structure not affect the outcomes of degree-based and betweenness-based strategies? This is left for future studies. In addition, selecting the initial blue nodes after considering all factors (using combined rank) can obtain even better results. A further study of such hybrid strategies can be a potential avenue for future research.

Author Contributions: Conceptualization, A.N.Z.; methodology, A.N.Z.; software, S.L.; formal analysis, S.L.; investigation, S.L. and A.N.Z.; resources, A.N.Z.; data curation, S.L.; writing—original draft preparation, S.L.; writing—review and editing, A.N.Z.; visualization, S.L.; supervision, A.N.Z.; project administration, A.N.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting this study are available on request.

Acknowledgments: We express our deepest appreciation to Serge Galam for his invaluable advice and feedback on this paper. This paper would not have been possible without his guidance and support.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Yin, X.; Wang, H.; Yin, P.; Zhu, H. Agent-based opinion formation modeling in social network: A perspective of social psychology. *Phys. A Stat. Mech. Its Appl.* **2019**, *532*, 121786. [[CrossRef](#)]
2. Bredereck, R.; Elkind, E. Manipulating opinion diffusion in social networks. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; Sierr, C., Ed.; International Joint Conference on Artificial Intelligence (IJCAI)/Information Sciences Institute: Marina del Rey, CA, USA, 2017; pp. 894–900. [[CrossRef](#)]
3. Huang, P.-Y.; Liu, H.-Y.; Chen, C.-H.; Cheng, P. The impact of social diversity and dynamic influence propagation for identifying influencers in social networks. In Proceedings of the WI-IAT'13: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 17–20 November 2013; IEEE Computer Society: NW Washington, DC, USA, 2013; Volume 1, pp. 410–416. [[CrossRef](#)]
4. Auletta, V.; Caragiannis, I.; Ferraioli, D.; Galdi, C.; Persiano, G. Minority becomes majority in social networks. In *Web and Internet Economics: Proceedings of the 11th International Conference WINE 2015*, Amsterdam, The Netherlands, 9–12 December 2015; Markakis, E., Schäfer, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 74–88. [[CrossRef](#)]
5. Auletta, V.; Ferraioli, D.; Greco, G. Reasoning about consensus when opinions diffuse through majority dynamics. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm and Twenty-Third European Conference on Artificial Intelligence (IJCAI-ECAI 2018), Sweden, 13–19 July 2018; Lang, J., Ed.; International Joint Conference on Artificial Intelligence (IJCAI)/Information Sciences Institute: Marina del Rey CA, USA, 2018; pp. 49–55. [[CrossRef](#)]
6. Galam, S. Minority opinion spreading in random geometry. *Eur. Phys. J. B* **2002**, *25*, 403–406. [[CrossRef](#)]
7. Galam, S. Heterogeneous beliefs, segregation, and extremism in the making of public opinions. *Phys. Rev. E* **2005**, *71*, 046123. [[CrossRef](#)] [[PubMed](#)]
8. Galam, S.; Jacobos, F. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Phys. A* **2007**, *381*, 366–376. [[CrossRef](#)]
9. Qian, S.; Liu, Y.; Galam, S. Activeness as a key to counter democratic balance. *Phys. A* **2015**, *432*, 187–196. [[CrossRef](#)]
10. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In KDD-2003: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; Domingos, P., Faloutsos, C., Senator, T., Kargupta, H., Getoor, L., Eds.; Association for Computing Machinery (ACM): New York, NY, USA, 2003; pp. 137–146. [[CrossRef](#)]
11. Talukder, A.; Alam, M.G.R.; Tran, N.H.; Niyato, D.; Park, G.H.; Hong, C.S. Threshold estimation models for linear threshold-based influential user mining in social networks. *IEEE Access* **2019**, *7*, 105441–105461. [[CrossRef](#)]
12. Zhuang, Z.; Wang, K.; Wang, J.; Zhang, H.; Wang, Z.; Gong, Z. Lifting majority to unanimity in opinion diffusion. In Proceedings of the ECAI 2020: 24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020; De Giacomo, G., Catala, A., Dilkina, B., Milano, M., Barro, S., Bugarin, A., Lang, J., Eds.; IOS Press: Amsterdam, The Netherlands, 2020; pp. 259–266. [[CrossRef](#)]
13. Avin, C.; Lotker, Z.; Mizrahi, A.; Peleg, D. Majority vote and monopolies in social networks. In ICDCN'19: Proceedings of the 20th International Conference on Distributed Computing and Networking, Bangalore, India, 4–7 January 2019; Hansdah, R.C., Krishnaswamy, D., Vaidya, N., Eds.; Association for Computing Machinery (ACM): New York NY, USA, 2019; pp. 342–351. [[CrossRef](#)]
14. Amir, G.; Baldasso, R.; Beilin, N. Majority dynamics and the median process: Connections, convergence and some new conjectures. *Stoch. Process Their Appl.* **2023**, *155*, 437–458. [[CrossRef](#)]
15. Anagnostopoulos, A.; Becchetti, L.; Cruciani, E.; Pasquale, F.; Rizzo, S. Biased opinion dynamics: When the devil is in the detail. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), Yokohama, Japan, 7–15 January 2021; Bessiere, C., Ed.; International Joint Conferences on Artificial Intelligence (IJCAI)/Information Sciences Institute: Marina del Rey CA, USA, 2021; pp. 53–59. [[CrossRef](#)]
16. Gekle, S.; Peliti, L.; Galam, S. Opinion dynamics in a three-choice system *Eur. Phys. J. B* **2005**, *45*, 569–575. [[CrossRef](#)]
17. Poljak, S.; Turzik, D. On pre-periods of discrete influence systems. *Discret. Appl. Math.* **1986**, *13*, 33–39. [[CrossRef](#)]

18. Zehmakan, A.N. Random majority opinion diffusion: Stabilization Time, absorbing states, and influential nodes. In Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, UK, 29 May 2023–2 June 2023; Ricci, A., Yeoh, W., Agmon, N., An, B., Eds.; International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS): Liverpool, UK, 2023; pp. 2179–2187. Available online: <https://www.ifaamas.org/Proceedings/aamas2023/forms/contents.htm#6E> (accessed on 7 November 2023).
19. Abdullah, M.A.; Draief, M. Global majority consensus by local majority polling on graphs of a given degree sequence. *Discret. Appl. Math.* **2020**, *180*, 1–10. [[CrossRef](#)]
20. Cruise, J.; Ganesh, A. Probabilistic consensus via polling and majority rules. *Queueing Syst.* **2014**, *78*, 99–120. [[CrossRef](#)]
21. Gärtner, B.; Zehmakan, A.N. Threshold behavior of democratic opinion dynamics. *J. Stat. Phys.* **2015**, *178*, 1442–1466. [[CrossRef](#)]
22. Auletta, V.; Ferraioli, D.; Grece, G. On the effectiveness of social proof recommendations in markets with multiple products. In Proceedings of the ECAI 2020: 24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020; De Giacomo, G., Catala, A., Dilkina, B., Milano, M., Barro, S., Bugarin, A., Lang, J., Eds.; IOS Press: Amsterdam, The Netherlands, 2020; pp. 19–26. [[CrossRef](#)]
23. Karia, N.; Mallick, F.; Dey, P. How hard is safe bribery? In Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, 9–13 May 2022; Faliszewski, P., Mascardi, V., Pelachaud, C., Taylor, M.E., Eds.; International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS): Liverpool, UK, 2022; pp. 714–722. Available online: <https://www.ifaamas.org/Proceedings/aamas2022/forms/contents.htm#1> (accessed on 7 November 2023).
24. Schoenebeck, G.; Tao, B.; Yu, F.-Y. Limitations of greed: Influence maximization in undirected networks re-visited. In Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, 9–13 May 2020; An, B., Yorke-Smith, N., El Fallah Seghrouchni, A., Sukthankar, G., Eds.; International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS): Liverpool, UK, 2020; pp. 1224–1232. Available online: <https://www.ifaamas.org/Proceedings/aamas2020/forms/contents.htm#RTP> (accessed on 7 November 2023).
25. Kundu, S.; Murthy, C.A.; Pal, S.K. A new centrality measure for influence maximization in social networks. In *Pattern Recognition and Machine Intelligence: Proceedings of the 4th International Conference PReMI 2011*, Moscow, Russia, 27 June–1 July 2011; Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 242–247. [[CrossRef](#)]
26. Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09), Paris, France, 28 June–1 July 2009; Elder, J, Soulié Fogelman, F., Flach, P., Zaki, M., Eds.; Association for Computing Machinery (ACM): New York, NY, USA, 2009; pp. 199–208. [[CrossRef](#)]
27. Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; Glance, N. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2007), San Jose, CA, USA, 12–15 August 2007; Berkhin, P., Caruana, R., Wu, X., Gaffney, S., Eds.; Association for Computing Machinery (ACM): New York, NY, USA, 2007; pp. 420–429. [[CrossRef](#)]
28. Goyal, A.; Lu, W.; Lakshmanan, L.V.S. CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In WWW'11: Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March 2011–1 April 2011; Sadagopan, S., Ramamritham, K., Kumar, A., Ravindra, M.P., Bertino, E., Kumar, R., Eds.; Association for Computing Machinery (ACM): New York, NY, USA, 2011; pp. 47–48. [[CrossRef](#)]
29. Wu, H.; Liu, W.; Yue, K.; Huang, W.; Yang, K. Maximizing the spread of competitive influence in a social network oriented to viral marketing. In *Web-Age Information Management: Proceedings of 16th International Conference WAIM 2015, Qingdao, China, 8–10 June 2015*; Dong, X., Yu, X., Li, J., Sun, Y., Eds.; Springer: Cham, Switzerland, 2015; pp. 516–519. [[CrossRef](#)]
30. Kamarthi, H.; Vijayan, P.; Wilder, B.; Ravindran, B.; Tambe, M. Influence maximization in unknown social networks: Learning policies for effective graph sampling. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2020), Auckland, New Zealand, 9–13 May 2020; An, B., Yorke-Smith, N., El Fallah Seghrouchni, A., Sukthankar, G., Eds.; International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS): Liverpool, UK, 2020; pp. 575–583. Available online: <https://www.ifaamas.org/Proceedings/aamas2020/forms/contents.htm#RTP> (accessed on 7 November 2023).
31. Zhao, G.; Jia, P.; Huang, C.; Zhou, A.; Fang, Y. A machine learning based framework for identifying influential nodes in complex networks. *IEEE Access* **2020**, *8*, 65462–65471. [[CrossRef](#)]
32. Li, Y.; Gao, H.; Gao, Y.; Guo, J.; Wu, W. A survey on influence maximization: From an ML-based combinatorial optimization. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–50. [[CrossRef](#)]
33. Leskovec, J.; SNAP: Stanford Large Network Dataset Collection. Available online: <http://snap.stanford.edu/data> (accessed on 7 November 2023).
34. Newman, M.E.J. *Networks: An Introduction*; Oxford University Press: Oxford, UK, 2018; Ch. 7. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.