# Information diffusion in complex networks: a model based on hypergraphs and its analysis



*Publisher:*

Springer

*Published version:*

DOI:10.1007/978-3-030-48478-1_3

# Information Diffusion in Complex Networks: a model based on hypergraphs and its analysis⋆

Alessia Antelmi[1], Gennaro Cordasco[2], Carmine Spagnuolo[1], and Przemysław Szufel[3]

[1] Dipartimento di Informatica, Università degli Studi di Salerno, Italy
aantelmi@unisa.it, cspagnuolo@unisa.it
[2] Dipartimento di Psicologia, Università degli Studi della Campania "Luigi Vanvitelli", Italy gennaro.cordasco@unicampania.it
[3] SGH Warsaw School of Economics, Poland pszufe@sgh.waw.pl

**Abstract.** This work introduces the problem of social influence diffusion in complex networks, where vertices are linked not only through simple pairwise relationships to other nodes but with groups of nodes of arbitrary size. A challenging problem that arises in this domain is to determine a small subset of nodes $S$ (*a target-set*) able to spread their influence in the whole network. This problem has been formalized and studied in different ways, and many viable solutions have been found for graphs. These have been applied to study several phenomena in research fields such as social, economic, biological, and physical sciences.
In this contribution, we investigated the *social influence problem* on hypergraphs. As hypergraphs are mathematical structures generalization of graphs, they can naturally model the many-to-many relationships characterizing a complex network. Given a network represented by a hypergraph $H = (V, E)$, we consider a dynamic influence diffusion process on $H$, evolving as follows. At the beginning of the process, the nodes in a given set $S \subseteq V$ are influenced. Then, at each iteration, the influenced hyperedges set is augmented by all hyperedges having a sufficiently large number of influenced nodes. Consequently, the set of influenced nodes is extended by all the nodes contained in a sufficiently large number of already influenced hyperedges. The process terminates when no new nodes can be influenced.
The so defined problem is an inherent chicken-and-egg question as nodes are influenced by groups of other nodes (or hyperedges), while hyperedges (or group of nodes) are influenced by the nodes they contain. In this paper, we provide a formal definition of the influence diffusion problem on hypergraphs. We propose a set of greedy-based heuristic strategies for finding the minimum influence target set, and we present an in-depth analysis of their performance on several classes of random hypergraphs. Furthermore, we describe an experiment on a real use-case, based on the character co-occurrences network of the Game-of-Thrones TV Series.

**Keywords:** Influence Diffusion · Target Set Selection· Random Hypergraphs· Social Network.

⋆ The research is financed by NAWA — The Polish National Agency for Academic Exchange.

## 1 Introduction

The current research on social networks is focusing on modeling community structures to examine how and to what extent relationships between humans or entities are the cause of complex emergent phenomena [40, 21]. In the past decades, graphs have played an essential role in the modeling and analysis of large-scale online social networks (OSNs) [8, 35], such as Facebook, Twitter or Linkedin, as well as for studying biological [26, 38] or economic systems [24, 41]. Adopting graphs to model these networks assumes the existence of only binary relationships between nodes. However, in many cases, complex networks are characterized by more articulated interactions. For instance, communication networks, reviewing activities, money transactions, geographical tracking, and many other scenarios are governed by many-to-many relationships. For a more clarifying example, we can consider the network built upon email exchanges between some users. In this context, the object *email* can be modeled as a relation involving a group of users. Thus, in this case, nodes of the network represent the persons, while the edges of the network incorporate a sub-set of them – i.e., all email receivers. It is worth noting that if we represent this scenario with a graph, we lose the information about which users are receivers of the same emails. This approach, combined with grouping messages having the same title, can be used for anomaly and spam detection in electronic communication [39]. Recently, hypergraphs have been exploited as a tool for modeling complex networks. Being a generalization of graphs, where a (hyper)edge is a relationship among an arbitrary number of nodes, they can naturally define many-to-many relations between groups of objects, such as domain names and IP addresses [30].

This research constitutes a relatively new area investigated in several recent works [16, 33]. A well-known problem in the field of network analysis is the question of social influence maximization, which aims to identify the set of nodes able to spread information in the whole network. However, little research on this topic does take into account many-to-many relationships existing in a complex network. Social influence [13, 18] is the process by which each individual change its behavior or adapt its opinions, according to the interactions with other people. With this aim in mind, it is crucial to notice that this process is a fundamental aspect in many fields, such as *viral marketing* [22, 11], in which the information diffusion process is used to attract people to adopt products or ideas. According to Lately [32], "*the traditional broadcast model of advertising-one-way, one-to-many, read-only is increasingly being superseded by a vision of marketing that wants and expects, consumers to spread the word themselves*". The major contributions of this paper are summarized as follows.

- We formally define the dynamic social influence problem on hypergraphs, and we present a variant of the target set problem, first presented in [31], suitable for networks involving many-to-many relationships.
- We introduce four random hypergraphs generative algorithms to build i) random hypergraphs (without any constraint); ii) $k$-uniform hypergraphs (where each hyperedge has size $k$); iii) $d$-regular hypergraphs (where each node has degree $d$); and iv) hypergraphs with the preferential attachment rule [5].

- We propose three greedy-based heuristics for finding the minimum influence target set on hypergraphs that eventually will influence the whole network.
- We present an evaluation of the proposed algorithms on a set of random hypergraphs, varying the random properties of the networks, and results on a real use-case, based on the network induced by the co-occurrences of characters in the Game-of-Thrones TV Series.

*Outline of the paper.* The paper is organized as follows. In Section 2, we define the minimum target set problem on hypergraphs, representing networks defined by many-to-many relationships. Furthermore, we describe four generating models of random hypergraphs. Section 3 reviews some relevant literature about the social influence problem and its applications. In Section 4, we describe our proposed greedy-based heuristics to solve the social influence problem. Section 5 presents our experiments, and we also discuss results on a real use-case. Finally, Section 6 details the conclusion and future work.

## 2 Background

### 2.1 Hypergraphs

A hypergraph is an ordered pair $H = (V, E)$ where $V$ is the set of nodes or vertices, which refers to a set of objects, and $E$ is the set of (hyper)edges. Each hyperedge is a non-empty subset of vertices; i.e., $E \subseteq 2^V \setminus \{\emptyset\}$, where $2^V$ is the power set of $V$. In this paper, we indicate with $n = |V|$ the number of nodes in $V$, and with $m = |E|$ the number of hyperedges in $E$, respectively. A graph is a hypergraph, where each hyperedge is a two element subset of $V$; in other words, a hypergraph $G = (V, E)$ is a graph if $E \subseteq \binom{V}{2} \subseteq 2^V \setminus \{\emptyset\}$. For a hypergraph $H$, a two-section representation $[H]_2$ can be obtained by connecting two nodes in the graph $[H]_2$ if and only if they are in the same hyperedge of $H$ [9]. As a result, each hyperedge from $H$ occurs as a complete graph in $[H]_2$. In this work, we considered the weighted $[H]_2$ of $H$, which assumes that the weight of an edge corresponds to the number of hyperedges containing both the edge endpoints.

### 2.2 Dynamic social influence diffusion on hypergraphs

Given a network represented by a hypergraph $H = (V, E)$, we consider a dynamic influence diffusion process on $H$, which evolves in discrete steps as follows. In the beginning, the nodes in a given set $S \subseteq V$ are influenced. Then, at each iteration:

1. the influenced hyperedges set is augmented by all edges which have a sufficiently large number of influenced nodes;
2. consequently, the set of influenced nodes is augmented by all the nodes which have a sufficiently large number of already influenced edges.

The process ends if no new nodes can be influenced.

Formally, let $H = (V, E)$ be a hypergraph. For each $v \in V$, we denote with $E(v) \subseteq E$ the set of edges that contains $v$ and with $d(v) = |E(v)|$ the degree of

$v$. Analogously, for each $e \in E$, we denote with $V(e) \subseteq V$ the set of nodes in $e$ and with $k(e) = |V(e)|$ the cardinality of $e$. Let $t_V : V \to \mathbb{N} = \{0, 1, \ldots\}$ and $t_E : E \to \mathbb{N} = \{0, 1, \ldots\}$ be two functions assigning thresholds to the vertices and to the hyperedges, respectively. For each node $v \in V$ (resp. $e \in E$), the value $t_V(v)$ (resp. $t_E(e)$) quantifies how hard it is to influence node $v$ (edge $e$), in the sense that easy-to-influence elements of the network have "low" threshold values, and hard-to-influence elements have "high" threshold values.

**Definition 1.** *Let $H = (V, E)$ be a hypergraph with threshold functions $t_V : V \longrightarrow \mathbb{N}$ and $t_E : E \longrightarrow \mathbb{N}$, and $S \subseteq V$. An* information diffusion process *in $H$, starting with a seed $S \subseteq V$, is a sequence*

$$I_V[S,0] \subseteq I_V[S,1] \subseteq \ldots \subseteq I_V[S,\ell] \subseteq \ldots \subseteq V$$

*of vertex subsets, with $I_V[S,0] = S$, and*

$$I_E[S,0] \subseteq I_E[S,1] \subseteq \ldots \subseteq I_E[S,\ell] \subseteq \ldots \subseteq E$$

*of edge subsets, with $I_E[S,0] = \emptyset$ and and such that for all $\ell > 0$*

$$I_E[S,\ell] = I_E[S,\ell-1] \bigcup \Big\{ e \in E : |V(e) \cap I_V[S,\ell-1]| \geq t_E(e) \Big\}$$

$$I_V[S,\ell] = I_V[S,\ell-1] \bigcup \Big\{ v \in V : |E(v) \cap I_E[S,\ell]| \geq t_V(v) \Big\}$$

*A **target set** for $H$ is a seed set $S \subseteq V$ that will eventually influence the whole network (i.e., $I_V[S,r] = V$ for some $r \geq 0$).*

We indicate the above information diffusion process on $H$ with

$$I_V[S], I_E[S] = \Phi(H, S, t_V, t_E),$$

where, $I_V[S] \subseteq V$ is the set of influenced vertices ($I_V[S] = I_V[S,r]$), and $I_E[S] \subseteq E$ is the set of influenced hyperedges. $t_V$ and $t_E$ denote the thresholds functions for nodes and hyperedges, respectively.

*Example 1.* Consider the hypergraph $H$ in Fig. 1. The nodes are depicted as an oval shape. The number on the top represents the node identifier; on the bottom, its threshold value is shown. The hyperedge threshold value is drawn as a black half oval shape. The hyperedge identifier is depicted inside the hyperedge. Finally, influenced nodes are drawn in gray. Influenced hyperedges are shaped using a gray dotted line. Given a possible seed set $S$ for $H$ equal to $\{v_1, v_4\}$, the information diffusion process evolves as follows.

$I_E[S,0] = \emptyset$, $I_V[S,0] = S = \{v_1, v_4\}$
$I_E[S,1] = \{e_2\}$, $I_V[S,1] = \{v_1, v_3, v_4\}$
$I_E[S,2] = \{e_2, e_3\}$, $I_V[S,2] = \{v_1, v_3, v_4, v_5\}$
$I_E[S,3] = \{e_1, e_2, e_3\}$, $I_V[S,3] = \{v_1, v_2, v_3, v_4, v_5\} = V$.
Hence, $S$ is a target set for $H$.

The problem examined in this paper is defined as follows:

*Problem 1.* **Diffusion on Hypergraphs — DoH**
**Instance:** $H = (V, E)$, thresholds $t_V : V \to N_0$ and $t_E : E \to N_0$.
**Problem:** Find a seed set $S \subseteq V$ of minimum size such that $I_V[S] = V$.

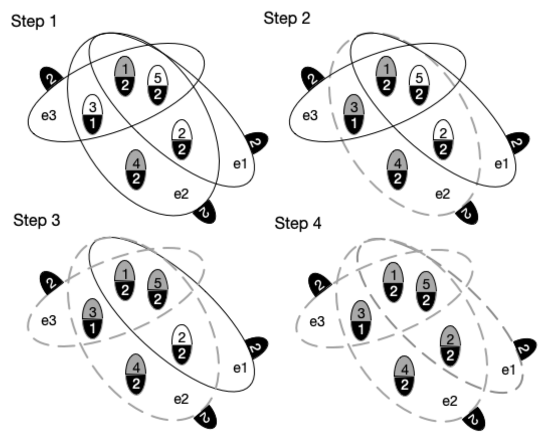



Fig. 1: An example of social influence diffusion process on $H = (V, E)$.

### 2.3 Models for random hypergraphs

In this work, we investigate information diffusion processes on complex networks by exploiting random hypergraphs. Here, we describe four generative models, characterized according to the structural proprieties of the computed random hypergraph (for example, hypergraphs with a fixed degree of nodes).

1. *Random model.* It generates a hypergraph without any structural property constraint. Given two integer parameters $n$ and $m$ (the number of nodes and hyperedges, respectively), the algorithm computes - for each hyperedge $he = 1, \ldots, m$ - a random number $s \in [1, n]$ (i.e. the hyperedge size). Then, the algorithm selects uniformly at random $s$ vertices from $V$ to add in $he$.
2. *K-uniform model.* It generates a $k$-uniform hypergraph, which is a hypergraph where each hyperedge has a size of $k$. The algorithm proceeds as the *random* model, but forcing the size of each hyperedge equal to $k$.
3. *D-regular model.* It generates a $d$-regular hypergraph, where each node has degree $d$. The algorithm exploits the $k$-uniform approach, described above, to build a $d$-regular hypergraph $H$ having $m$ nodes and $n$ edges. It then returns the hypergraph $H^*$, dual of $H$.
4. *Preferential-attachment model.* It generates a hypergraph with a preferential attachment rule between nodes, as described in [5]. The algorithm starts with a fully-random graph with 5 nodes and 5 edges. It then iteratively adds a node or an edge, according to a given parameter $p$, defining the probability of creating a new node or a new hyperedge. In detail, the connections with the new node or hyperedge are generated according to a preferential attachment policy [5]. We slightly changed the algorithm to avoid repetitions in the hyperedges.

## 3 Related Work

***The social influence diffusion problem.*** Previous research showed the importance of the target set selection (TSS) problem to study the social influence diffusion in networks. The TSS problem aims to select $k$ initially-influenced seed users to maximize the expected number of eventually-influenced users. In other words, the objective is to find a subset of nodes in the network that, once active, can activate all the nodes of the network under the linear threshold (LT) influence propagation model. According to the LT model, a user $v$ becomes active when the sum of influences of its neighbors in the networks reaches a specific threshold $t(v)$ [17]. Given its importance in the context of influence spread in both the online (social networks) and offline (word-of-mouth) worlds, the TSS problem is extensively studied on graphs. Kempe et al. [31] first analyzed the problem in networks with randomly chosen thresholds. Chen [12] studied the minimization problem of finding the smallest target set able to influence the whole network built with fixed arbitrary thresholds. Furthermore, Chen proved a strong inapproximability result that makes unlikely the existence of an algorithm for the TSS problem on graphs (2-uniform hypergraphs) with an approximation factor better than $O(2^{\log^{1-\epsilon}|V|})$. Cordasco et al. [16] presented an algorithm for the TSS always producing an optimal solution (i.e., a minimum size subset of nodes that influence the whole network) in case the network is either a tree, a cycle, or a complete graph.

Considering that researchers started focusing on hypergraphs only in the last decade, little or no literature exists on the TSS problem on hypergraphs. Zhu et al. [42] deal with the problem of social influence maximization in social networks. They model the crowd influence as a hyperedge $e = (H_e, v)$ with weight $0 \leq P_e \leq 1$, where $H_e$ is the head node-set and $v$ is the tail node, meaning that $v$ will be activated by $H_e$ with probability $P_e$ only after each node in $H_e$ is activated. Their proposed algorithm selects $k$ initially-influenced seed users in a directed hypergraph $G = (V, E, P)$, maximizing the expected number of eventually-influenced users. Another stochastic diffusion process in which information diffusion can occur through interactions in groups of different sizes is described by Iacopini et al. [28].

Our study addresses the social influence diffusion problem on networks characterized by many-to-many relationships, using undirected hypergraphs, which allow modeling more kinds of real-world use cases, such as social networks like Facebook or Yelp. Furthermore, in our work, we adopted a linear thresholds model, investigating different threshold values for nodes and hyperedges. We also present a deterministic model which is more suitable for real use-cases.

***Random hypergraphs generation.*** The foundations of random graph theory lie in a seminal paper by Erdős and Rényi [20]. However, several models have been developed that make it possible to generate random graphs having desired topological properties to better mimic the real world. The Barabási-Albert models *rich-get-richer* phenomena. On the other hand Watts-Strogat *small-world* model is useful for representation of social networks. Random graph structures have proved to be a useful concept in many disciplines. Still, more complex mathematical tools are needed to comprehensively and accurately model many real-world complex networks [5]. The study of random hypergraph models has its origin from work by Erdős and Bollobas [7], which presents an analogous to the Erdős-Rényi random graph model. In the following years, researchers focused on analyzing several properties of this model [15, 19, 23, 1]. Wang et al. [29] first defined a preferential attachment model for hypergraphs, with vertex arrival events and constant-size hyperedges. Starting from this model and its limitations, Avin et al. [5] proposed a preferential attachment model generating hypergraphs with hyperedges of arbitrary size, allowing cycles and non-uniformity. In particular, they extended the Chung-Lu preferential attachment model proposed for graphs [14].

## 4 Finding the Minimum Target Set on Hypergraphs

In this Section, we discuss three greedy-based heuristics for the *DoH* problem (see Section 2.2), i.e., finding the minimum influence target set $S \subseteq V$ of a hypergraph $H = (V, E)$ able to influence the whole network. A simple greedy strategy may be selecting - at each iteration - the nodes in descending order by their degree until the current set can influence the whole network. We refer to this approach with the label *StaticGreedy*. It enables us to compute the set $S$ by exploiting a binary search strategy detailed in Algorithm 1. As described in Section 2.2, we indicate the diffusion process on $H$ with $\Phi(H, S)$, and we denote with $I_V[S] \subseteq V$ and $I_E[S] \subseteq E$ the end set of influenced nodes and hyperedges, respectively.

**Algorithm 1** *StaticGreedy*$(H = (V, E), t_V, t_E)$

Let $\sigma(V)$ be the list of nodes in descending order of their degree $d(v)$.
$left = 1$, $right = |V|$
**while** $left < right$ **do** ▷ Binary Search
  $mid = \left\lceil \frac{left+right}{2} \right\rceil$
  $I_V[S], I_E[S] = \Phi(H, \sigma_{mid}, t_V, t_E)$ ▷ $\sigma_i$ denotes the set containing the first $i$ nodes in the order $\sigma(V)$;
  **if** $I_V[S] != V$ **then**
    $left = mid$
  **else**
    $right = mid - 1$
return $S = \sigma_{left+1}$

A dynamic approach, referred to as *DynamicGreedy*, is listed in Algorithm 2. In this heuristic, all nodes are added to the candidates set $U$. At each stage, the node of the maximum degree is added to $S$ and removed from $U$. At this point, some nodes and/or hyperedges become infected. The algorithm simulates the diffusion process, and influenced edges are pruned from the network.The degree of nodes ($\delta(v)$) is updated accordingly.

**Algorithm 2** *DynamicGreedy*$(H = (V, E), t_V, t_E)$

$S = \emptyset$, $U = V$, $E' = E$
**for** $u \in U$ **do**
  $\delta(u) = d(u)$
**while** $U \neq \emptyset$ **do**
  $v = argmax_{u \in U} \delta(v)$
  $U = U \setminus \{v\}$
  $S = S \cup \{v\}$
  $I_V[S], I_E[S] = \Phi(H, S, t_V, t_E)$
  **if** $I_V[S] = V$ **then**
    break;
  $E' = E - I_E[S]$
  **for** $u \in U$ **do**
    $\delta(u) = |E(u) \cap E'|$ ▷ $\delta(u)$ denotes the degree of $u$ in $H = (V, E')$.
return $S$

Given the *DynamicGreedy* algorithm, we have designed a similar heuristic, named *DynamicGreedy*$_{[H]_2}$, and listed in Algorithm 2. In this heuristic, we compute the degree of the nodes on the $[H]_2$ of the residual hypergraph $H^i$ of $H$. $H^i$ is the hypergraph obtained removing all hyperedges already influenced by the nodes in $S$ at stage $i$.

**Algorithm 3** *DynamicGreedy*$_{[H]_2}(H(V, E), t_V, t_E)$

$S = \emptyset$, $U = V$, $E' = E$, $[H]_2 = 2Section(H(V, E))$
**while** $U \neq \emptyset$ **do**
  $v = argmax_{u \in U} d_{[H]_2}(v)$ ▷ $d_{[H]_2}(v)$ denotes the degree of $v$ in $[H]_2$.
  $U = U \setminus \{v\}$
  $S = S \cup \{v\}$
  $I_V[S], I_E[S] = \Phi(H, S, t_V, t_E)$
  **if** $I_V[S] = V$ **then**
    break;
  $E' = E - I_E[S]$
  $[H]_2 = 2Section(H(V, E'))$
return $S$

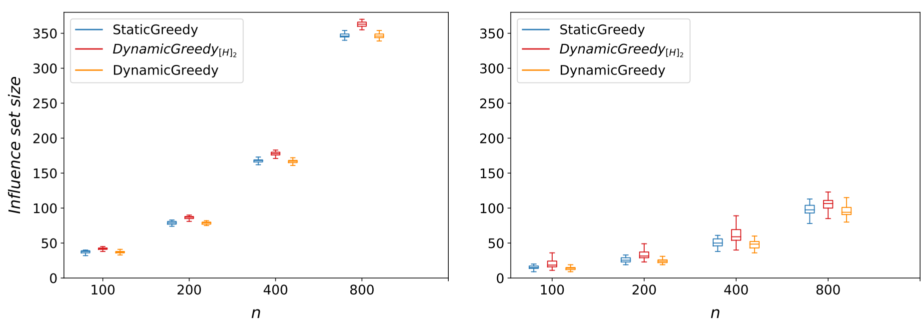
## 5 Experiments

We present experiments on the three greedy-based heuristics discussed in Section 4. We investigated two classes of experiments; we evaluated the proposed heuristics on random networks, and on a real use-case by exploiting the co-occurrences network of the TV Series Game-of-Thrones.

### 5.1 Random networks

We performed three experimental scenarios for the case of random hypergraphs. In the first and second scenarios, we fixed the node threshold to a random value between 1 and its degree. In the last scenario, each node threshold varies proportionally - from 0.1 to 0.9 - to the degree of the node. In particular, in the first scenario, we run the heuristics on random hypergraphs with no structural proprieties generated with the *random* model and hypergraphs generated with the *preferential-attachment* rule. We ranged the hypergraph size, using $[100, 200, 400, 800]$ nodes and hyperedges. In the second scenario, we experimented the heuristics on *k-uniform* and *d-regular* random hypergraphs, ranging the value of $k$ and $d$ in $[10, 20, 40, 80]$. In the third and last scenario, we generated a random hypergraph of fixed size ($n = m = 500$) with all generative models. We fixed both $k = 80$ and $d = 80$, for the *k-uniform* and *d-regular* random hypergraphs, respectively. In all experiments, we set each hyperedge activation threshold proportional to its degree scaled of factor 0.5 (majority policy). We executed each experiment 48 times. We implement all heuristics and experiments in *Julia* language, by exploiting the library `SimpleHypergraphs.jl` [3]. The Julia code used in the paper is available at the following public GitHub repository[4].

*Scenario 1 — **Increasing $H$ size, random thresholds.*** Fig. 2 shows the results obtained on random hypergraphs - with different sizes - generated by the *random* and *preferential-attachment* models. On the $y$-axis, we report the size of the influence target set S; on the $x$-axis, the hypergraph size ($n = m$). The *DynamicGreedy* heuristic achieves the best average performance. However, as shown in Fig. 2a, there is not a significant difference between the three strategies. On the other hand, the *DynamicGreedy* heuristic significantly outperforms the others in the case of the *preferential-attachment* scenario, as shown in Fig. 2b.



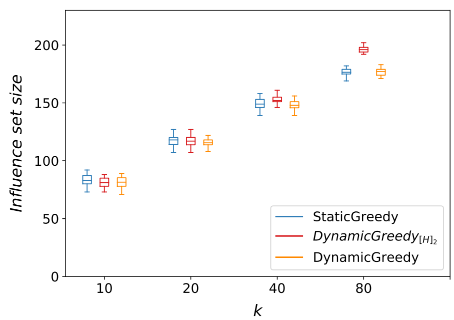

(a) *Random model.* (b) *Preferential-attachment model.*

Fig. 2: Experiments on random hypergraphs $H = (V, E)$, generated with the random and preferential-attachment models, varying the degree of nodes and hyperedges ($n = m$). For each node, the threshold is fixed to a random value between 1 and the node degree. A fixed threshold to 0.5 is used for hyperedges.

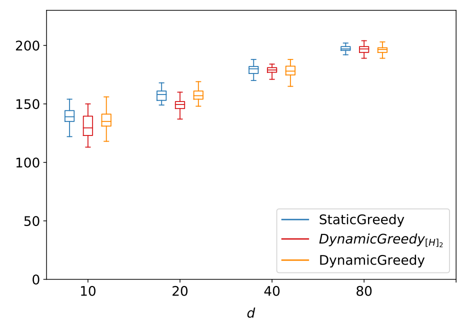
*Scenario 2 — **Uniform and Regular $H$, random thresholds.*** Fig. 3 shows the results obtained on random hypergraphs generated by the *k-uniform* and *d*-

[4] https://github.com/pszufe/LTMSim.jl

*regular* models. On the $x$-axis, we show values for $k$ and $d$. As shown in Fig. 3a, the *DynamicGreedy* strategy achieves better results for random $k$*-uniform* hypergraphs, especially in the case of large values of $k$. Fig. 3b depicts the results for random $d$*-regular* hypergraphs. By increasing the size of $d$, there is no significant difference between the heuristics, even if for small values of $d$, their results exhibit a more significant variance. It is worth discussing the interesting - even though not so surprising - outcomes revealed by the comparison of the results obtained in the $k$*-uniform* and $d$*-regular* experiments. In general, the $k$*-uniform* hypergraphs require a target set of smaller size compared to $d$*-regular* hypergraphs.



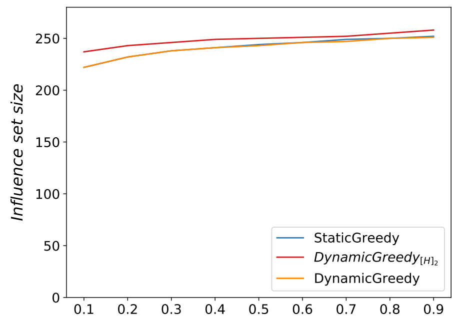

(a) *K-uniform model.*



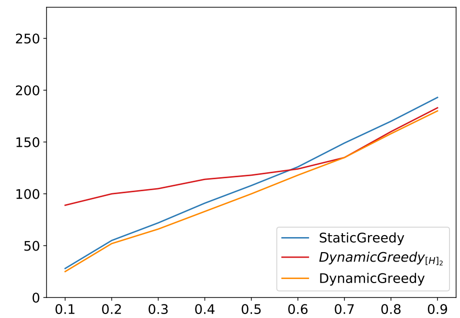

(b) *D-regular model.*

Fig. 3: Experiments for random $k$-uniform and $d$-regular hypergraphs $H = (V, E)$, with a fixed hypergraph size $n = m = 500$. For each node, the threshold is fixed to a random value between 1 and the node degree. A fixed threshold to 0.5 is used for hyperedges.

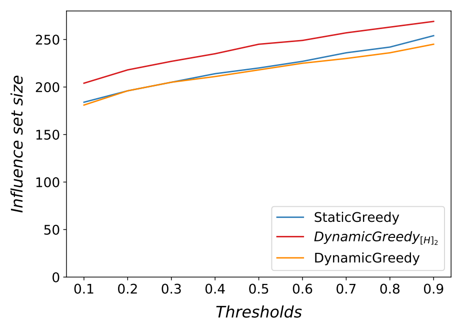
*Scenario 3* —***Varying node thresholds proportionally to their degree.*** Fig. 4 outlines the results obtained on a hypergraph $H$ of fixed size $n = m = 500$, generated by each random model. We ranged nodes activation thresholds proportionally to their degree size from 0.1 to 0.9, and we fixed the hyperedges activation threshold proportionally to 0.5. The heuristics achieve almost the same performance in the case of a completely *random* graph (Fig. 4a) and a $d$*-regular* (Fig. 4d) hypergraph. Results obtained from the *preferential-attachment* (Fig. 4b) and $k$*-uniform* (Fig. 4c) models are more attractive. In both experiments, $DynamicGreedy_{[H]_2}$ exhibits the worst results compared to the other two heuristics. Interestingly, the preferential-attachment case exhibits unusual behavior. When the thresholds are small, the performance of $DynamicGreedy_{[H]_2}$ is poor, but for larger values, its performance improves and is very close to the *DynamicGreedy* heuristic. As a result of using high threshold values, it is hard to trigger an information cascade in the network as, in this case, the influence



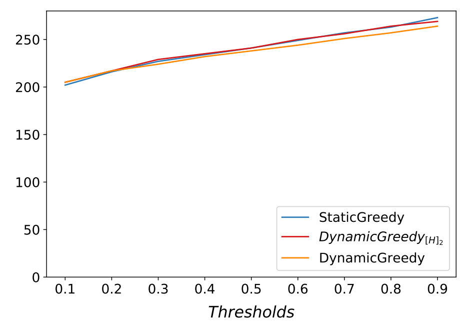

(a) *Random model.*




(b) *Preferential-attachment model.*




(c) $80_k$*-uniform model.*




(d) $80_d$*-regular model.*

Fig. 4: Experiments for random hypergraphs $H = (V, E)$, of size $n = m = 500$, considering all random generative models. For each node, the threshold varies proportionally - from 0.1 to 0.9 - to the degree of the node. A fixed threshold to 0.5 is used for hyperedges. The value of $k$ and $d$ for the $k$-uniform and $d$-regular hypergraphs is set to 80.

diffusion process behaves more like a domination process. In general, this makes the problem easier to face.

### 5.2 Game-of-Thrones TV series network

Game of Thrones [25] (GoT) is the screen adaption of the series of fantasy novels *A Song of Ice and Fire*, written by George R.R. Martin. Created by D. Benioff and D.B. Weiss for the American television network HBO, the American fantasy drama TV series has attracted a record viewership and has a broad, active, and international fan base—according to Wikipedia [5]. This enthusiasm has led the intricate world of GoT to be a profoundly immersive entertainment experience [4]. Both the academic community and industries took the opportunity to study not only complex dynamics within the GoT storyline [6], but also how

[5] https://en.wikipedia.org/wiki/Game_of_Thrones

viewers engage with the GoT world on social media [2, 27, 37], or how the novel itself is a portrait of real-world dynamics [10, 34, 36].
In this experiment, we exploited GoT season episodes data from the dataset *Game of Thrones Datasets and Visualizations*, available at the following GitHub repository[6]. Specifically, we used information describing each episode scenes. They contain - for each scene - start, end, location, and a list of characters performing in it. Table 1 reports some necessary information about the number of episodes, scenes, and characters per GoT season. A more detailed description of the dataset is available on the dataset GitHub repository.

***The GoT network* — $H_{GoT}$.** We modeled the GoT network using a hypergraph $H_{got}$, considering the characters co-occurrences within scenes per each season. The vertices of $H_{got}$ represents the 577 GoT characters. Each hyperedge of $H_{got}$, therefore, link together all characters that have acted in the same scene together. The total number of considered scenes was 4165. Fig. 5 presents the hyperedges size distribution of $H_{got}$. It shows a typical power-law distribution, where few scenes assemble a considerable number of characters. In contrast, many others focus on few or no characters.

| Season | Episodes | Scenes | Characters |
|---|---|---|---|
| 1 | 10 | 286 | 125 |
| 2 | 10 | 468 | 137 |
| 3 | 10 | 470 | 137 |
| 4 | 10 | 517 | 152 |
| 5 | 10 | 508 | 175 |
| 6 | 10 | 577 | 208 |
| 7 | 7 | 468 | 75 |
| 8 | 6 | 871 | 66 |

Table 1: Some GoT dataset numbers.

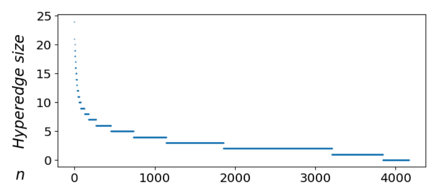


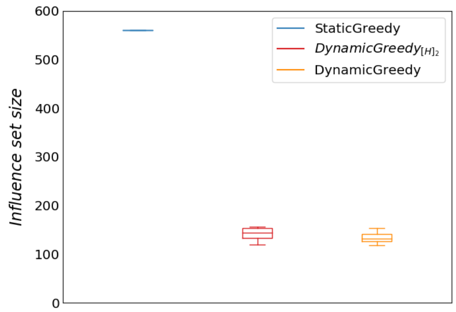

Fig. 5: $H_{GoT}$ hyperedges distribution.

***Influencing the GoT network.*** We performed two experiments on the GoT network, aiming at evaluating the performance of the heuristics in minimizing the number of nodes (or characters) to influence. In Fig. 6, we detail the performance of each heuristic both in the case of random threshold values for each node (Fig. 6a), and in the case of proportional threshold values (Fig. 6a). The *DynamicGreedy* and $DynamicGreedy_{[H]_2}$ provide similar results requiring a seed set of about 120 nodes on average. The second case shows the same trend, and they can find reasonable solutions and achieve, in the worst-case (0.9), a target set of size about 30% of $V$. On the contrary, *StaticGreedy* provides a target set almost equal to $V$ for each threshold.

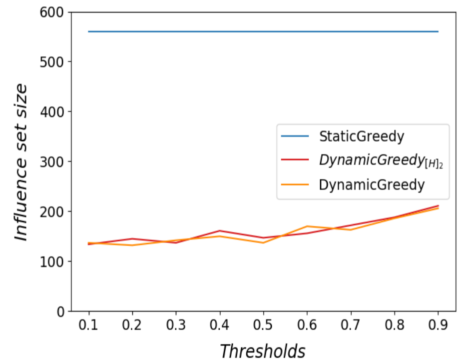
## 6 Conclusions and Future Work

This paper faces the social influence diffusion process in complex networks, exploiting the hypergraph structure. We propose a formulation of the dynamic

[6] Game of Thrones Datasets and Visualizations.
https://github.com/jeffreylancaster/game-of-thrones by Jeffrey Lancaster




(a) Random thresholds.




(b) Degree proportional thresholds.

Fig. 6: Experiments for the GoT network using a) random and b) proportional nodes thresholds values. A fixed threshold to 0.5 is used for hyperedges.

influence diffusion on hypergraphs, referred to as the *Diffusion on Hypergraphs* (DoH) problem. The so-defined problem on hypergraphs differs from the correspondent on graphs, as it introduces the influence propagation also on the network connections, i.e., the hyperedges (which denote groups of related nodes).

A challenging problem arising in this domain is to determine a small subset of nodes $S$ (*a target-set*) able to spread their influence in the whole network. We present three greedy-based heuristics to solve this problem on hypergraphs, considering either the degree of nodes in the hypergraph $H$ or in the two-section view $[H]_2$ of $H$, and selecting the nodes according to static or dynamic policies. We provided an exhaustive investigation of their performance on a bunch of random networks and a real use-case based on the character co-occurrences in the GoT TV series. We observed that the *DynamicGreedy* heuristic achieved the best results in the case of random networks. In the real use-case of the GoT network, experiments highlighted that dynamically selecting the nodes (according to their degree in the residual hypergraph) to add to the target set results in a more efficient solution compared to a static approach. Furthermore, for the GoT network, we also noticed that the dynamic greedy-based heuristics (*DynamicGreedy* and $DynamicGreedy_{[H]_2}$) provided a good seed set when choosing an initial set of size at most 30% of $V$.

As future work, we plan to investigate more efficient algorithms and approaches for the *DoH* problem. Furthermore, we aim to experiments with the proposed strategies on real-world datasets, such as a *Twitter* social network built upon tweet hashtags or user reviews from the *Yelp.com* dataset. Results are encouraging, and further investigation is needed to explore the social influence diffusion problem on hypergraphs as it might shed light on complex social phenomena, like fake news sharing in online social networks.

## References


1. Aksoy, S. G., Kolda, T. G., and Pinar, A. Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks 5*, 4 (03 2017), 581–603.
2. Antelmi, A., Breslin, J., and Young, K. Understanding user engagement with entertainment media: A case study of the twitter behaviour of game of thrones (got) fans. In *2018 IEEE Games, Entertainment, Media Conference (GEM)* (2018).
3. Antelmi, A., Cordasco, G., Kamiński, B., Prałat, P., Scarano, V., Spagnuolo, C., and Szufel, P. Simplehypergraphs.jl—novel software framework for modelling and analysis of hypergraphs. In *Algorithms and Models for the Web Graph* (2019), pp. 115–129.
4. Askwith, I. D. Television 2.0 : reconceptualizing TV as an engagement medium. https://dspace.mit.edu/handle/1721.1/41243, 2007.
5. Avin, C., Lotker, Z., and Peleg, D. Random preferential attachment hypergraphs. *Computer Science 23* (2015).
6. Beveridge, A., and Shan, J. Network of thrones. *Math Horizons 23*, 4 (2016).
7. Bollobas, B., and Erdös, P. Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society 80*, 3 (1976), 419–427.
8. Bonato, A., Eikmeier, N., Gleich, D. F., and Malik, R. Dynamic competition networks: Detecting alliances and leaders. In *Algorithms and Models for the Web Graph* (2018).
9. Bretto, A. *Hypergraph Theory: An Introduction*. 2013.
10. Bruno, L., and Miriam, H. The house of black and white: Identities of color and power relations in the game of thrones. *Revista Nós 4*, 2 (2019), 161–182.
11. Cautis, B., Maniu, S., and Tziortziotis, N. Adaptive influence maximization. In *Procs of the 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining* (2019), p. 3185–3186.
12. Chen, N. On the approximability of influence in social networks. *SIAM Journal on Discrete Mathematics 23*, 3 (2009), 1400–1415.
13. Chen, W., Lakshmanan, L. V. S., and Castillo, C. *Information and Influence Propagation in Social Networks*. Morgan & Claypool Publishers, 2013.
14. Chung, F. R. K., and Lu, L. Complex graphs and networks. In *CBMS Regional Conference Series in Mathematics* (2006).
15. Cooper, C., Frieze, A., Molloy, M., and Reed, B. Perfect matchings in random r-regular, s-uniform hypergraphs. *Combinatorics, Probability and Computing 5*, 1 (1996), 1–14.
16. Cordasco, G., Gargano, L., Mecchia, M., Rescigno, A. A., and Vaccaro, U. Discovering small target sets in social networks: A fast and effective algorithm. *Algorithmica 80*, 6 (2018), 1804–1833.
17. Cordasco, G., Gargano, L., and Rescigno, A. On finding small sets that influence large networks. *Social Network Analysis and Mining Vol 6(1)* (2016).
18. Easley, D., and Kleinberg, J. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
19. Ellis, D., and Linial, N. On regular hypergraphs of high girth. *Electronic Journal of Combinatorics 21*, 1 (2014).
20. Erdős, P., and Rényi, A. On the evolution of random graphs. In *Publication of the Mathematical Institute of Hungarian Academy Of Science* (1960), pp. 17–61.
21. Fani, H., Jiang, E., Bagheri, E., Al-Obeidat, F., Du, W., and Kargar, M. User community detection via embedding of social network structure and temporal content. *Information Processing and Management 57*, 2 (2020).

22. Ghayoori, A., and Nagi, R. Seed investment bounds for viral marketing under generalized diffusion. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2019), p. 95–100.
23. Ghoshal, G., Zlatic, V., Caldarelli, G., and E. J. Newman, M. Random hypergraphs and their applications. *Physical review. E, Statistical, nonlinear, and soft matter physics* (2009).
24. Göbel, M., and Araújo, T. A network structure analysis of economic crises. In *Complex Net. and Their App. VIII* (2020), pp. 547–560.
25. HBO. Game of Thrones. https://www.hbo.com/game-of-thrones, 2019.
26. Hossain, M., Khan, A., and Uddin, S. Understanding the progression of congestive heart failure of type 2 diabetes patient using disease network and hospital claim data. In *Complex Net. and Their App. VIII* (2020).
27. Héctor, J. P., and Rainer, R. On jon snow's death: Plot twist and global fandom in game of thrones. *Culture & Psychology* (2019).
28. Iacopini, I., Petri, G., Barrat, A., and Latora, V. Simplicial models of social contagion. *Nature Communications 10* (2019), 2485.
29. Jian-Wei, W., Li-Li, R., Qiu-Hong, R., and Ji-Yong, Z. Evolving hypernetwork model. *Physics of Condensed Matter 77* (10 2010), 493–498.
30. Joslyn, C., Aksoy, S., Arendt, D., Jenkins, L., Praggastis, B., Purvine, E., and Zalewski, M. High performance hypergraph analytics of domain name system relationships. In *HICSS Symp. on Cybersecurity Big Data Analytics* (2019).
31. Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2003), p. 137–146.
32. Lately, D. An Army of Eyeballs: The Rise of the Advertisee. https://thebaffler.com/latest/an-army-of-eyeballs, 2014. [Online; 2019].
33. Liqing, Q., Jinfeng, Y., Xin, F., Wei, J., and Wenwen, G. Analysis of influence maximization in temporal social networks. *IEEE Access 7* (2019).
34. Milkoreit, M. Pop-cultural mobilization: Deploying game of thrones to shift us climate change politics. *International Journal of Politics, Culture, and Society 32*, 1 (2019), 61–82.
35. Moutidis, I., and Williams, H. T. P. Utilizing complex networks for event detection in heterogeneous high-volume news streams. In *Complex Net. and Their App. VIII* (2020), pp. 659–672.
36. Muno, W. *"Winter Is Coming?" Game of Thrones and Realist Thinking.* 2019.
37. Rhodes, R. E., and Zehr, E. P. Fight, flight or finished: forced fitness behaviours in game of thrones. *British Journal of Sports Medicine 53*, 9 (2019), 576–580.
38. Romero, M., Finke, J., Quimbaya, M., and Rocha, C. In-silico gene annotation prediction using the co-expression network structure. In *Complex Net. and Their App. VIII* (2020).
39. Silva, J., and Willett, R. Hypergraph-based anomaly detection of high-dimensional co-occurrences. *IEEE transactions on pattern analysis and machine intelligence 31*, 3 (2008), 563–569.
40. Tien, J., Eisenberg, M., Cherng, S., and Porter, M. Online reactions to the 2017 'unite the right' rally in charlottesville: measuring polarization in twitter networks using media followership. *Applied Network Science 5*, 1 (2020).
41. Verba, M. A. "learning hubs" on the global innovation network. In *Complex Net. and Their App. VIII* (2020), pp. 620–632.
42. Zhu, J., Zhu, J., Ghosh, J., Wu, W., and Yuan, J. Social influence maximization in hypergraph in social networks. *IEEE Transactions on Network Science and Engineering 6*, 4 (2019), 801–811.