Stochastic pretopology as a tool for complex networks analysis

Quang Vu Bui\textsuperscript{a,b}, Soufian Ben Amor\textsuperscript{c} and Marc Bui\textsuperscript{a,d}

\textsuperscript{a}CHArt Laboratory EA 4004, EPHE, PSL Research University, Paris, France; \textsuperscript{b}CHArt Laboratory EA, University of Sciences, Hue, Hue, Vietnam; \textsuperscript{c}LI-PARAD Laboratory, University of Versailles-Saint-Quentin-en-Yvelines, Paris-Saclay University, Paris, France; \textsuperscript{d}University Paris 8, Paris, France

ABSTRACT

We are proposing in this work a novel approach for complex network analysis by introducing Stochastic Pretopology, a result of the combination of Pretopology theory and Random Sets. We firstly show how pretopology generalizes the graph theory to deal with complex interactions in complex networks. By connecting with random set theory, we then give the definition of stochastic pretopology and also propose the ways to build this kind of pretopology in many situations. That is to say how proximity with randomness can be delivered to model complex neighbourhoods formation in complex networks. In addition, we also show how stochastic pretopology can be applied for modelling dynamic processes on complex networks by representing classical information diffusion models under stochastic pretopology language and then proposing Pretopology Cascade Model as a general model for information diffusion process that can take place in more complex networks such as multi-relational networks or stochastic graphs.

ARTICLE HISTORY

Received 30 June 2018
Accepted 28 October 2018

KEYWORDS
Random set; pretopology; complex network; multi-relational network; information diffusion

1. Introduction

Complex system is a system composed of many interacting parts, such that the collective behaviour of its parts together is more than the ‘sum’ of their individual behaviours (Newman, 2011). The topology of complex systems (who interact with whom) is often specified in terms of complex networks that are usually modelled by graphs, composed by vertices or nodes and edges or links. Graph theory has been widely used in the conceptual framework of network models, such as random graphs, small world networks, scale-free networks (Easley & Kleinberg, 2010; Newman, 2003).

However, having more complicated non-regular topologies, complex networks need a more general framework for their representation (Newman, 2011). To overcome this issue, we propose using Stochastic Pre-Topology built from the mixing between Pretopology theory (Belman, 2011) and Random Sets theory (Molchanov, 2005; Nguyen, 2006). Pre-Topo-logy is a mathematical tool for modelling the concept of proximity which allows us to
follow structural transformation processes as they evolve while random sets theory provides the good ways for handling what happens in a stochastic framework at the sets’ level point of views.

In this paper, after recalling basic concepts of pretopology and the definition of graph in the *Claude Berge sense* (Berge, 1962), we first show pretopology as an extension of graph theory which leads us to the definition of pretopology networks as a general framework for network representation and then present an example of a complex group interactions model by using pretopology theory.

By connecting to random sets theory for dealing with the random factors that often occur in real-world complex networks, we then give the definition of *Stochastic Pretopology* and show how we construct stochastic pseudo-closure functions for various contexts. These functions are useful for modelling the complex networks in different spaces such as metric space, valued or binary relation spaces, etc. These models can be also convenient to handle phenomena in which collective behaviour of a group of elements can be different from the summation of element behaviours composing the group.

After showing how stochastic pretopology can be applied for modelling dynamic processes on complex networks by representing classical information diffusion models such as *Independent Cascade model* (Goldenberg, Libai, & Muller, 2001) and *Independent Threshold model* (Granovetter, 1978) under stochastic pretopology language, we then propose *Pretopology Cascade Model* as a general information diffusion model in which complex random neighbourhoods set can be captured by using stochastic pseudo-closure functions. We also present this model in two specific kinds of complex networks: multi-relational networks and stochastic graphs. Stochastic graphs presented in this paper are defined by extending the definition of graph in the *Berge sense* (Berge, 1962) \( G = (V, \Gamma) \). In this approach, by considering \( \Gamma \) function as a finite random set defined from a degree distribution, we give a general graph-based network model in which *Erdős–Rényi model* and *scale-free networks* are special cases. For proposed models, after describing the algorithms, we illustrate the model by running some experiments.

The rest of this paper is organized as follows: Section 2 briefly recalls basic concepts of pretopology theory; Section 3 shows pretopology as a tool for network modelling, while Section 4 present stochastic pretopology when we connect pretopology and random set. After representing some classical information diffusion models in the Sections 5 and 6 presents *Pretopology Cascade Model* as an application of stochastic pretopology in information diffusion and we then conclude our work in Section 7. The work in this paper is an extended version of our paper presented in ACIIDS 2018 conference (Bui, Amor & Bui, 2018) by adding approximate 60% new material: algorithms for proposal models, some experiments illustrated theoretical models and some improvements in the background parts.

### 2. Pretopology theory

For modelling the dynamic processes on complex networks, topology theory is not suitable since the idempotent property of its closure function makes it impossible for changing unless changing the topological structure. So, we propose in this section a new insight on networks modelling with pretopology theory. Pretopology (Belmandt, 2011) is considered as an extension of topology obtained by the relaxing of its axiomatic. The
pretopology is a tool for modelling the concept of proximity that allows monitoring step by step the evolution of a set. It establishes a powerful tool for the structure analysis, classification (Belmandt, 2011), and multi-criteria clustering (Bui, Sayadi, & Bui, 2016). It also applies for group modelling in social networks (Levorato, 2011). Based on set theory, by considering a group of elements as a set, pretopology formalism allows us to consider a group as a whole independent entity.

2.1. Pretopological notions

**Definition 2.1.** A pretopological space is an ordered pair \((X, a)\), where \(X\) is a set and \(a: \mathcal{P}(X) \rightarrow \mathcal{P}(X)\) is a **pseudo-closure** operator, satisfying the two following axioms:

- \((P1): a(\emptyset) = \emptyset; \) (Preservation of Nullary Union)
- \((P2): A \subseteq a(A) \ \forall \ A, A \subset \ X \) (Extensivity)

It is important to note that, by defining **pseudo-closure** \(a(.)\) (see Figure 1(a)) we do not suppose that it is an idempotent transform. Then, conversely as it happens in topology, we can compute: \(a(A), a(a(A)), a(a(a(A))), \ldots, a^k(A)\) (see Figure 1(b)). So, **pseudo-closure** allows, for each of its applications, to add elements to a set departure according to defined characteristics. The starting set gets bigger but never reduces.

**Definition 2.2.** Let \((X, a)\) a pretopological space, \(\forall A, A \subset X\). \(A\) is a closed subset if and only if \(a(A) = A\).

**Definition 2.3.** Given a pretopological space \((X, a)\), call the closure of \(A\), when it exists, the smallest closed subset of \(X\) which contains \(A\). The closure of \(A\) is denoted by \(F(A)\).

Closure is very important because of the information it gives about the influence or reachability of a set, meaning, for example, that a set \(A\) can influence or reach elements into \(F(A)\), but not further (see Figure 1(b)).

Hence, it is necessary to build a pretopological spaces in which the closure always exists. We present in the following different types of pretopological spaces which are

![Figure 1](https://via.placeholder.com/150)

**Figure 1.** Pseudo-closure and closure function. (a) Pseudo-closure of \(A\). (b) Closure of \(A\).
less general than the basic ones but for which, good properties are fulfilled by neighbourhoods as well as allowing the existence of closure.

2.2. Pretopological spaces

**Definition 2.4.** A Pretopology space \((X, a)\) is called \(\mathcal{V}\)-type space if and only if

\[(P3) \quad (A \subseteq B) \Rightarrow (a(A) \subseteq a(B)) \quad \forall A, B \in \mathcal{P}(X) \quad \text{(Isotonic)} \quad (1)\]

**Definition 2.5.** A Pretopology space \((X, a)\) is called \(\mathcal{V}_D\)-type space if and only if

\[(P4) \quad a(A \cup B) = a(A) \cup a(B) \quad \forall A \subseteq X, \quad \forall B \subseteq X \quad \text{(Additive)} \quad (2)\]

**Definition 2.6.** A Pretopology space \((X, a)\) is called \(\mathcal{V}_S\)-type space if and only if

\[(P5) \quad a(A) = \bigcup_{x \in A} a(\{x\}) \quad \forall A \subseteq E \quad (3)\]

2.3. Some ways to build pseudo-closure function

2.3.1. Pretopology in metric space

Let us consider space \(X\) is endowed with a metric defined by a distance \(d\). Let \(r\) be a positive real. For each element \(x\) of \(X\), \(B(x, r)\) is a ball with the center \(x\) and a radius \(r\) defined by

\[B(x, r) = \{y \in X \mid d(x, y) \leq r\}\]

Clearly, \(B(x) = (B(x, r)_{x \in E})\) is basis of neighbourhoods of \(X\). We then can build a pseudo-closure function \(a(.)\) on \(X\) with \(B(x, r)\) such as:

\[\forall A \in \mathcal{P}(X), \quad a(A) = \{x \in X \mid B(x, r) \cap A \neq \emptyset\} \quad (4)\]

Pretopological build in previous is of \(\mathcal{V}_D\)-type pretopological space (\(\mathcal{V}_S\)-type if \(X\) is finite). The pseudo-closure \(a(A)\) is a set of all elements \(y \in X\) such that \(y\) is within a distance of at most radius \(r\) from at least one element of \(A\) (see Figure 2(a) for an example).

2.3.2. The elements of \(X\) are equipped with a neighbourhood function

Let us consider a multivalued function \(\Gamma : X \rightarrow \mathcal{P}(X)\) as a neighbourhood function. \(\Gamma(x)\) is a set of neighbourhoods of element \(x\). We define a pseudo-closure \(a(.)\) (see Figure 2 for an example) as follows:

\[\forall A \in \mathcal{P}(X), \quad a(A) = A \cup \left(\bigcup_{x \in A} \Gamma(x)\right) \quad (5)\]

Clearly, pretopology space built from pseudo-closure \(a(.)\) function defined in previous is of \(\mathcal{V}_S\) type. Graph is defined by Claude Berge sense (Berge, 1962) is a special case of this kind of pretopology space.
2.3.3. The elements of \( X \) are linked by \( n \) reflexive binary relations

Suppose we have a family \((R_i)_{i=1}^n\) of binary reflexive relations on a finite set \( X \). For each relation \( R_i \), we can define a pretopological structure by considering the following subset:

\[
\forall i = 1, 2, \ldots, n, \forall x \in X, V_i(x) \text{ defined by:}
\]

\[
V_i(x) = \{ y \in X | x R_i y \}
\]

Then, the pseudo-closure \( a_i(\cdot) \) is defined by:

\[
a_i(A) = \{ x \in X | \forall i = 1, 2, \ldots, n, V_i(x) \cap A \neq \emptyset \} \quad \forall A \subseteq X
\]

Pretopology defined on \( X \) by \( a_i(\cdot) \) using the intersection operator is called the strong pretopology. Figure 3(a) gives an example for strong pretopology built from two relationships.

**Proposition 2.7.** \( a_i(\cdot) \) determines on \( X \) a pretopological structure and the spaces \((X, a_i)\) is of \( V_i\)-type pretopological space.

Similarly, we can define weak pretopology from \( a_w(\cdot) \) by using the union operator:

\[
a_w(A) = \{ x \in X | \exists i = 1, 2, \ldots, n, V_i(x) \cap A \neq \emptyset \} \quad \forall A \subseteq X
\]

**Proposition 2.8.** \( a_w(\cdot) \) determines on \( X \) a pretopological structure and the space \((X, a_i)\) is of \( V_i\)-type.

2.3.4. The elements of \( X \) are linked by a valued relation

In order to model certain problems such as model in weighted graph, we often need the space \( X \) are bound by a valued relation. For instance, we can define a real value \( \nu \) on relations as a function from \( X \times X \rightarrow \mathbb{R} \) as: \((x, y) \rightarrow \nu(x, y)\). The pseudo-closure \( a(\cdot) \) can build such as:

\[
\forall A \in \mathcal{P}(X), \quad a(A) = \left\{ y \in X - A \left\lfloor \sum_{x \in A} \nu(x, y) \geq s \right\} \cup A; s \in \mathbb{R} \right\}
\]

The pseudo-closure \( a(\cdot) \) is composed of \( A \) and of all elements \( y \) where the sum of valued edges between some elements of \( A \) and \( y \) is greater than the threshold \( s \). Pretopological

---

**Figure 2.** Pseudo-closure in metric space and a space equipped with a neighbour function. (a) Pseudo-closure of \( A \) in a metric space. (b) Pseudo-closure of \( A \) in a space equipped with a neighbour function.
space built from this kind of pseudo-closure is \( V \)-type space. Figure 3(b) gives an illustration of this space with \( s = 10 \).

3. Pretopology as a tool for network modelling

3.1. Pretopology as an extension of Graph theory

3.1.1. Graphs in the Berge sense

By using the knowledge from multivalued function, Claude Berge (Berge, 1962) defined a graph such as:

**Definition 3.1.** A graph, which is denoted by \( G = (V, \Gamma) \), is a pair consisting of a set \( V \) of vertices or nodes and a multivalued function \( \Gamma \) mapping \( V \) into \( P(V) \).

The pair \((x, y)\), with \( y \in \Gamma(x) \) is called an arc or edge of the graph. We therefore can also denote a graph by a pair \( G = (V, E) \), which \( V \) is a set of nodes and \( E \) is a set of edges. Conversely, if we denote a graph as \( G = (V, E) \), we can define the \( \Gamma \) function as: \( \Gamma(x) = \{ y \in V \mid (x, y) \in E \} \). \( \Gamma(x) \) is a set of neighbours of node \( x \).

3.1.2. Pretopology as an extension of Graph theory

In this part, we show reflexive graph \((V, \Gamma)\) which is a special case of pretopology. More specifically, as it is known, a finite reflexive graph \((V, \Gamma)\) complies the property: \( \forall A \subseteq V, a(A) = \bigcup_{x \in A} \Gamma(x) \) where pseudo-closure function defined as \( a(A) = \bigcup_{x \in A} \Gamma(x) \). For this reason, graph may be represented by a \( V \)-type pretopological space. Conversely, we can build a pretopology space \((V, a)\) presented a graph such as: \( a(A) = \{ x \in V \mid \Gamma(x) \cap A \neq \emptyset \} \) where \( \Gamma(x) = \{ y \in V \mid x \, R \, y \} \) built from a binary relation \( R \) on \( V \). Therefore, a graph \((V, \Gamma)\) is a pretopological space \((V, a)\) in which the pseudo-closure function built from a binary relation (see Section 2.3.3) or built from a neighbourhood function in Equation (5).

By using a graph, a network is represented by only one binary relation. In the real world, however, a network is a structure made of nodes that are tied by one or more specific
types of binary or value relations. As we show in the previous, by using pretopology theory, we can generalize the definition of complex network such as:

**Definition 3.2. ( Pretopology network).** A pretopology network, which is denoted by $G^{(Pretopo)} = (V, a)$, is a pair consisting of a set $V$ of vertices and a pseudo-closure function $a(.)$ mapping $\mathcal{P}(V)$ into $\mathcal{P}(V)$.

### 3.2. Example of a complex group interactions model

In this section, we show an example in which we can deal with complex interactions in social networks modelling. We use the *Sampson Monastery* dataset\(^1\) from the work of Sampson (1968) which consists of social relations among a set of 18 monk-novitiates preparing to enter a monastery and has been often used in sociology studies. The data include a variety of relations, such as esteem, liking, dislike, influence, praise and so on. A lot of relations are coded, however, in this example, we concentrate on two of them: **like** (Figure 4(a)) and **dislike** (Figure 4(b)). Since each member ranked only his top three choices on that tie, relations are nonsymmetric and weighted with three integer values, from 1 to 3 (3 indicates the highest or first choice and 1 the last choice).

In our model, for group coalitions, we consider the question: ‘who wants to join the group?’. Based on the two relations **like** and **dislike**, we can see that: people who have the greatest like for B should join his group more than the others but this person will not accept in the group people he does not like, even if these people have like for him. The question here is how we can model this kind of social networks.

To overcome this aim, we firstly set two valued relations as functions $\text{like}(x, y)$, $\text{dislike}(x, y)$ from $X \times X$ to $\mathbb{N}$ with $X, \text{like, dislike}$ being the set containing individuals, *like* relation, *dislike* relation respectively and then propose the way a new member $x$ joins to a group $A \in \mathcal{P}(X)$ with two steps:

**ST1:** if person $x$ has **like** relations with some members in group $A$ and the sum of all **like** weights is greater than a chosen threshold $\theta_1$ then he will send a message ‘**Want to join**’ to this group.

![Figure 4. Sampson Networks with two relations like and dislike. (a) Sampson Like Network. (b) Sampson DisLike Network.](image-url)
ST2: when group A receive a ‘Want to join’ message from person x, the group will see if the current members of the group have any ‘problem’ with x by considering the dislike relation. If sum of all dislike weights from all member of the group to x is less than a chosen threshold $\theta_2$, they will accept this member and send a message ‘Accept for joining’.

We therefore can build the group coalitions model of $A \in P(X)$ by defining pseudo-closure $a(A)$ as:

$$a(A) = \left\{ y \in X : \sum_{x \in A} like(y, x) > \theta_1 \land \sum_{x \in A} dislike(x, y) < \theta_2 \right\}; \quad \theta_1, \theta_2 \in \mathbb{N} \quad (9)$$

By setting up the values of two parameters $\theta_1$, $\theta_2$, we can capture strong or weak restriction for growing a group. If we apply this pseudo-closure on a set A, people would be in $a(A)$ if they have like relation with one or several persons of A (according to $\theta_1$) and if no people or a few people of A (according to $\theta_2$) have dislike relation with them. Figure 5 shows an example with $\theta_1 = 3$, $\theta_2 = 2$. Figure 5(a) shows an illustration of pseudo-closure function of a set $A=[\text{HUGH}_14, \text{JOHN}_1]$. We can see WINF_12 will join to group A since he has like relations with JOHN_1 (weight =3) and HUGH_14 (weight =1) and nobody in this group has any ‘problem’ with him. The same situation occurs with BONI_15 but it is different with GREG_2 since JOHN_1 dislikes to him with weight = 2 although GREG_2 would like to join this group. VICTOR_8 does not want to joint the group since he just likes only JOHN_1 with weight=3 and even if he wants to joint the group in the future, the group does not accept since HUGH_14 dislikes to him with weight=2.

The pretopology space built in this example is neither $V_D$-type nor $V_S$-type since $a([1]) \cup a([14]) = [1] \cup [14]$ while $a([1, 14]) = [1, 12, 14, 15]$ with $1, 12, 14, 15$ denote JOHN_1, WINF_12, HUGH_14, BONI_15, respectively.

Figure 5. Example of the complex group interactions model with $\theta_1 = 3$, $\theta_2 = 2$. Like relation is plain and dislike relation is dashed. (a) Pseudo-closure of A. (b) Largest group started with two members found according to the neighbourhood definition.
In this model, the question of finding the largest group started with 2 members following the rules of our neighbourhood can be treated by building closure function. When the closure function is applied to all subsets of \(X\) which contains two elements, it reveals that \([\text{GREG}_2, \text{BONAVEN}_5]\) is the group that can form the largest amount of people in the network (Figure 5(b)).

4. Stochastic pretopology (SP)

Complex systems usually involve structural phenomena, under stochastic or uncontrolled factors. In order to follow these phenomena step by step, we need concepts which allow modelling dynamics of their structure and take into account the factors’ effects. As we showed in the previous section, we propose to use pretopology for modelling the dynamics of phenomena; the non idempotents of its pseudo-closure function makes it suitable for such a modelling. Then, we introduce stochastic aspects to handle the effects of factors influencing the phenomena. For that, we propose using a theory of random sets by considering that, given a subset \(A\) of the space, its pseudo-closure \(a(A)\) is considered as a random set. So, we have to consider the pseudo-closure not only as a set transform but also as a random correspondence.

Stochastic pretopology was first basically introduced in chapter 4 of (Belmandt, 2011) by using a special case of random set (the simple random set) to give three ways to define stochastic pretopology. We have also given some applications of stochastic pretopology such as: modelling pollution phenomena (Lamure, Bonnevay, Bui, & Amor, 2009) or studying complex networks via a stochastic pseudo-closure function defined from a family of random relations (Basileu, Amor, Bui, & Lamure, 2012). Since we will deal with complex networks in which set of nodes \(V\) is a finite set, we propose in this paper another approach for building stochastic pretopology by using finite random set theory (Nguyen, 2006).

From now on, \(V\) denotes a finite set. \((\Omega, \mathcal{A}, P)\) will be a probability space, where: \(\Omega\) is a set, representing the sample space of the experiment; \(\mathcal{A}\) is a \(\sigma\)-algebra on \(\Omega\), representing events and \(P: \Omega \rightarrow [0,1]\) is a probability measure.

4.1. Finite random set

**Definition 4.1.** A finite random set (FRS) with values in \(\mathcal{P}(V)\) is a map \(Y: \Omega \rightarrow \mathcal{P}(V)\) such as

\[
Y^{-1}(A) = \{\omega \in \Omega: Y(\omega) = A\} \in \mathcal{A} \text{ for any } A \in \mathcal{P}(V)
\]  

(10)

The condition (10) is often called measurability condition. So, in other words, a FRS is a measurable map from the given probability space \((\Omega, \mathcal{A}, P)\) to \(\mathcal{P}(V)\), equipped with a \(\sigma\)-algebra on \(\mathcal{P}(V)\). We often choose \(\sigma\)-algebra on \(\mathcal{P}(V)\) is the discrete \(\sigma\)-algebra \(\mathcal{E} = \mathcal{P}(\mathcal{P}(V))\). Clearly, a finite random set \(Y\) is a random element when we refer to the measurable space \((\mathcal{P}(V), \mathcal{E})\). This is because \(Y^{-1}(A) \subseteq \mathcal{A}\) since \(\forall \mathcal{A} \in \mathcal{E}; Y^{-1}(\mathcal{A}) = \cup \mathcal{A} \in \mathcal{A} Y^{-1}(A)\).

4.2. Definition of stochastic pretopology

**Definition 4.2.** We define stochastic pseudo-closure defined on \(\Omega \times V\), any function \(a(\cdot, \cdot)\) from \(\Omega \times \mathcal{P}(V)\) into \(\mathcal{P}(V)\) such as:
\[(\Omega \times V, a(\cdot, \cdot))\] is then called Stochastic Pretopological space.

By connecting to the finite random set theory (Molchanov, 2005; Nguyen, 2006), we can build stochastic pseudo-closure function with different ways.

**4.3. SP defined from random variables in metric space**

By considering a random ball \(B(x, \xi)\) with \(\xi\) is a non-negative random variable, we can build a stochastic pseudo-closure \(a(\cdot)\) in metric space such as:

\[
\forall A \in \mathcal{P}(V), \quad a(A) = \{x \in V \mid B(x, \xi) \cap A \neq \emptyset\} \quad (11)
\]

**4.4. SP defined from random variables in valued space**

We present two ways to build stochastic pseudo-closure by extending the definition of pseudo-closure function on valued space presented in Equation (8). Firstly, by considering threshold \(s\) is a random variable \(\eta\), we can define a stochastic pseudo-closure \(a(\cdot)\) such as:

\[
\forall A \in \mathcal{P}(V), \quad a(A) = \left\{ y \in V - A \left| \sum_{x \in A} v(x, y) \geq \eta \right. \right\} \cup A \quad (12)
\]

where threshold \(\eta\) is random variable.

Secondly, by considering the weight function \(v(x, y)\) between two elements \(x, y\) as a random variable, we can define a stochastic pseudo-closure \(a(\cdot)\) such as:

\[
\forall A \in \mathcal{P}(V), \quad a(A) = \left\{ y \in V - A \left| \sum_{x \in A} v(x, y) \geq s \right. \right\} \cup A \quad (13)
\]

where \(v(x, y)\) is a random variable.

**4.5. SP defined from a random relation built from a family of binary relations**

Suppose we have a family \((R_i)_{i=1,\ldots,m}\) of binary reflexive relations on a finite set \(V\). We call \(L = \{R_1, R_2, \ldots, R_m\}\) is a set of relations. Let us define a random relation \(R: \Omega \rightarrow L\) as a random variable:

\[
P(R(\omega) = R_i) = p_i; \quad p_i \geq 0; \quad \sum_{i=1}^{m} p_i = 1.
\]

For each \(x \in V\), we can build a random set of neighbours of \(x\) with random relation \(R\):

\[
\Gamma_{R(\omega)}(x) = \{y \in V \mid x R(\omega) y\}
\]

We can define a stochastic pseudo-closure \(a(\cdot, \cdot)\) such as:

\[
\forall A \in \mathcal{P}(V), \quad a(\omega, A) = \{x \in V \mid \Gamma_{R(\omega)}(x) \cap A \neq \emptyset\} \quad (14)
\]
4.6. SP defined from a family of random relations

We can extend the previous work by considering many random relations. Suppose we have a family \((R_i)_{i=1,...,n}\) of random binary reflexive relations on a set \(V\). For each \(x \in V\), we can build a random set of neighbours of \(x\) with random relation \(R_i\), \(i = 1,2,\ldots,n\):

\[
\Gamma_{R_i(\omega)}(x) = \{y \in V \mid x R_i(\omega) y\}
\]

We can define a stochastic pseudo-closure \(a(\cdot,\cdot)\) such as:

\[
\forall A \in \mathcal{P}(V), \quad a(\omega,A) = \{x \in V \mid \forall i = 1,2,\ldots,n, \Gamma_{R_i(\omega)}(x) \cap A \neq \emptyset\} \quad (15)
\]

4.7. SP defined from a random neighbourhood function

Let us consider a random neighbourhood function as a random set \(\Gamma: \Omega \times V \to \mathcal{P}(V)\). \(\Gamma(\omega,x)\) is a random set of neighbourhoods of element \(x\). We define a stochastic pseudo-closure \(a(\cdot,\cdot)\) as follows:

\[
\forall A \in \mathcal{P}(V), \quad a(\omega,A) = A \cup \left( \bigcup_{x \in A} \Gamma(\omega,x) \right) \quad (16)
\]

We have shown in this section how we construct stochastic pseudo-closure functions for various contexts. That is to say how proximity with randomness can be delivered to model complex neighbourhoods formation in complex networks. In the two next sections, we will show how stochastic pretopology can be applied for modelling dynamic processes on complex networks by representing classical information diffusion models under stochastic pretopology language and then proposing Pretopology Cascade Model as a general information diffusion model in which complex random neighbourhoods set can be captured by using stochastic pseudo-closure functions.

5. Stochastic pretopology as a general information diffusion model on single relational networks

Information diffusion has been widely studied in networks, aiming to model the spread of information among objects when they are connected with each other. In a single relational network, many diffusion models have been proposed such as SIR model, tipping model, threshold models, cascade models, … (Easley & Kleinberg, 2010; Newman, 2003). Please refer the work of (Guille, Hacid, Favre, & Zighed, 2013) for a survey of information diffusion models in online social networks. In this section, we focus on independent cascade (IC) model (Goldenberg et al., 2001) which is a generalized of SIR model, and another model is known as linear threshold (LT) model (Granovetter, 1978). After describing properties of these models, we present in the following two scenarios in which stochastic pretopology as extensions of both IC model (Goldenberg et al., 2001) and IT model (Granovetter, 1978).
5.1. The independent cascade and linear threshold models

5.1.1. Model definitions

We assume a network $G = (V, \Gamma, W)$, where:

- $V$ is a set of vertices.
- $\Gamma : V \rightarrow \mathcal{P}(V)$: neighbourhood function.
  - $\Gamma(x)$ is set of outcoming neighbourhoods of node $x$.
  - $\Gamma^{-1}(x)$ is set of incoming neighbourhoods of node $x$.
- $W : V \times V \rightarrow \mathbb{R}$ is weight function.
  - in IT model, $W(x, y)$ is the weight of edge between two nodes $x, y$.
  - in IC model, $W(x, y)$ is the probability of node $y$ infected from node $x$.

The diffusion process occurs in discrete time steps $t$. If a node adopts a new behaviour or idea, it becomes active, otherwise it is inactive. An inactive node has the ability to become active. The set of active nodes, newly active nodes at time $t$ is considered as $A_t$, $A_{t}^{new}$, respectively. The tendency of an inactive node $x$ to become active is positively correlated with the number of its active incoming neighbours $\Gamma^{-1}(x)$. Also, we assume that each node can only switch from inactive state to active state, and an active node will remain active for the rest of the diffusion process. In general, we start with an initial seed set $A_0$ and through the diffusion process, for a given inactive node $x$, its active neighbours attempt to activate it. The process runs until no more activations occur.

5.1.2. Independent cascade model

In IC model, there is a probability of infection associated with each edge. $W(x, y)$ is the probability of node $x$ infecting $y$. This probability can be assigned based on the frequency of interactions, geographic proximity, or historical infection traces, ... Each node, once infected, has the ability to infect its neighbour in the next time step based on the probability associated with that edge. At each time step $t$, each node $x \in A_{t-1}^{new}$ infects the inactive neighbours $y \in \Gamma(x)$ with a probability $W(x, y)$. The propagation continues until no more infection can occur (see Algorithm 1).

\begin{algorithm}
\begin{algorithmic}[1]
\Require Network $G = (V, \Gamma, W)$, seed set $A_0$
\Procedure{Random-IC-Model}{$G, A_0$}
\State $t \leftarrow 0$; $A_{total} \leftarrow A_0$; $A_{new} \leftarrow A_0$
\While{infection occur}$t \leftarrow t + 1$; $A_t \leftarrow \emptyset$
\For{$u \in A_{new}$}
\State $A_t(u) \leftarrow \{v_{inactive} \in \Gamma(u), q \leq W(u, v_{inactive})\}; q \sim U(0, 1)$
\State $A_t \leftarrow A_t \cup A_t(u)$
\EndFor
\State $A_{total} \leftarrow A_{total} \cup A_t$; $A_{new} \leftarrow A_t$
\EndWhile
\State \Return $A_{total}$ \Comment{Output}
\EndProcedure
\end{algorithmic}
\caption{Independent Cascade model}
\end{algorithm}
5.1.3. Independent threshold model

In IT model, each directed edge \((x, y)\) has a non-negative weight \(W(x, y)\). For any node \(y \in V\), the total incoming edge weights sum to less than or equal to one, i.e. \(\sum_{x \in G(y) \cap A_{t-1}} W(x, y) \leq 1\). An active node influences its inactive neighbours according to the weights. At each step, an inactive node \(y\) becomes active if the total weight of its incoming neighbours is at least threshold \(\theta_y\), \(\theta_y \in [0, 1]\). The dynamics of the model are specified below.

Under the IT model, each node \(y\) selects a threshold \(\theta_y\) in the interval \([0, 1]\) uniformly at random. Then, at each time step \(t\) where \(A_{t-1}\) is the set of nodes activated at time \(t-1\) or earlier, each inactive node \(y\) becomes active if \(\sum_{x \in G(y) \cap A_{t-1}} W(x, y) \geq \theta_y\). The propagation continues until no more infection can occur (see Algorithm 2).

5.2. Stochastic pretopology as an extension of IC model

We can represent IC model by giving a definition of stochastic pretopology based on two definitions in Sections 4.4 and 4.7: we firstly define a random set of actived nodes from each node \(x \in A_{t-1}^{\text{new}}\) and then use a random neighbour function to define the random active nodes in the time \(t\). Specifically, \(A_t^{\text{new}}\) is defined via two steps: i. For each \(x \in A_{t-1}^{\text{new}}\), set of actived nodes from \(x\), \(\Gamma^{(\text{active})}(x)\), defined as:

\[
\Gamma^{(\text{active})}(x) = \{ y \in \Gamma(x) \mid W(x, y) \geq \eta \}; \quad \eta \sim U(0,1)
\]  

\hspace{3cm} (17)

ii. The set of newly active nodes, \(A_t^{\text{new}}\), defined as:

\[
A_t^{\text{new}} = a(A_{t-1}^{\text{new}}) - A_{t-1}; \quad A_t = A_{t-1} \cup a(A_{t-1}^{\text{new}})
\]  

\hspace{3cm} (18)

where

\[
a(A_{t-1}^{\text{new}}) = A_{t-1}^{\text{new}} \bigcup \left( \bigcup_{x \in A_{t-1}^{\text{new}}} \Gamma^{(\text{active})}(x) \right)
\]  

\hspace{3cm} (19)
5.3. Stochastic pretopology as an extension of IT model

We can represent IC model by giving a definition of stochastic pretopology such as:

$$A_t = a(A_{t-1}) = \left\{ y \in V - A_{t-1} \bigg| \sum_{x \in \Gamma^{-1}(y) \cap A_{t-1}} W(x, y) \geq \eta \right\} \cup A_{t-1}; \quad \eta \sim U(0,1)$$

6. Pretopology cascade models for modelling information diffusion on complex networks

6.1. Pretopology cascade model (PCM)

Most of the information diffusion models are defined via node’s neighbours. In general, at each time step $t$, the diffusion process can be described in two steps:

Step 1: define set of neighbours $N(A_{t-1})$ of set of active nodes $A_{t-1}$.

Step 2: each element $x \in N(A_{t-1}) - A_{t-1}$ will be influenced by all elements in $A_{t-1}$ to be active or not active node by following a diffusion rule.

We consider the way to define set of neighbours $N(A_{t-1})$ in step 1. In classical diffusion model with complex network represented by a graph $G = (V, \Gamma)$, $N(A_{t-1})$ is often defined such as: $N(A_{t-1}) = \bigcup_{x \in A_{t-1}, \Gamma(x)}$. By using the concepts of stochastic pretopology theory introduced in the Section 4, the information diffusion process can be generalized by defining a set of neighbours $N(A_{t-1})$ as a stochastic pseudo-closure function $N(A_{t-1}) = a_0(A_{t-1})$. We therefore propose the Pretopology Cascade Model presented in the following as a general information diffusion model which can be captured more complex random neighbourhoods set in diffusion processes.

**Definition 6.1.** Pretopology Cascade model:

Under the Pretopology Cascade model (see Algorithm 3), at each time step $t$, the diffusion process takes place in two steps:

Step 1: define set of neighbours $N(A_{t-1})$ of $A_{t-1}$ as a stochastic pseudo-closure function $N(A_{t-1}) = a_0(A_{t-1})$.

Step 2: each element $x \in N(A_{t-1}) - A_{t-1}$ will be influenced by $A_{t-1}$ to be active or not active node by following a ‘diffusion rule’.

---

**Algorithm 3** Pretopology Cascade model

**Require**: Set of elements $V$, stochastic pseudo-closure $a_0$, seed set $A_0$

```
1: procedure PRETOPO-CASCADE-MODEL(V, a_0, A_0)
2:      t ← 0, A^{total} ← A_0
3:      while infection occur do
4:         t ← t + 1; A_t ← ∅
5:         N_t ← a_0(A^{total}) \> Compute random neighbors using a_0
6:         for u ∈ N_t − A^{total} do
7:            if diffusion condition satisfied then
8:               A_t.append(u)
9:                 end if
10:        end for
11:      A^{total} ← A^{total} ∪ A_t \> Output
12:   end while
13:   return A^{total}
14: end procedure
```
For defining $N(A_{t−1})$ in step 1, we can apply different ways to define stochastic pseudo-closure function presented in Section 4. ‘Diffusion rule’ in step 2 can be chosen by various ways such as:

- Probability based rule: element $x$ infects the inactive elements $y \in N(A_{t−1})$ with a probability $P_{x,y}$.
- Threshold rule: inactive elements $y \in N(A_{t−1})$ will be activated if sum of all influence of all incoming elements of $y$ greater than a threshold $\theta_y$.

Discussion: why Pretopology Cascade Model is a general model for cascade model? Firstly, as we showed in the section 5, Independent Cascade Model and Linear Threshold model are specific cases of Pretopology Cascade Model. Secondly, by using pseudoclosure to define set of neighbours, we can capture more complex kind of neighbours of a set since we showed in the Section 3, ‘Pretopology Network’ is a general concept for Graph theory and we also proposed in the section 2 some ways to build pseudoclosure function in many situations for application. Finally, by using stochastic-pseudoclosure to define set of neighbours, we can also capture random factors often occur in Complex Networks.

We present in the following two specific cases of the PCM that can be useful in real-world situation: the first takes place in a stochastic graph by defining random neighbours sets based on nodes’ degree distribution and the second takes place in multi-relational networks where random neighbours set is built from a family of relations.

6.2. PCM on stochastic graphs

6.2.1. Stochastic graph

Definition 6.2. Stochastic Graph. A stochastic graph, which is denoted by $G^\Omega = (V, \Gamma^\Omega)$ is a pair consisting of a set $V$ of vertices and a finite random set $\Gamma^\Omega$ mapping $\Omega \times V$ into $P(V)$.

The random neighbour function $\Gamma^\Omega$ in the Definition 6.2 can be defined in a general way from finite random set theory (Nguyen, 2006). Since the nodes’ degree distribution is necessary for studying network structure, we propose here the way to defining the random neighbour function $\Gamma^\Omega$ via two steps:

1. Defining probability law of the cardinality of $\Gamma^\Omega$ (in fact, $\Gamma^\Omega$ is a degree distribution of network).

   \[
   \text{Prob}(|\Gamma^\Omega| = k) = p_k \quad \text{for } k = 1, 2, \ldots, \infty
   \]

2. Assigning probability law on $V^k$ for $k = 1, 2, \ldots, \infty$

   \[
   \text{Prob}(\Gamma^\Omega_1 = x^{(1)}, \ldots, \Gamma^\Omega_k = x^{(k)} | |\Gamma^\Omega| = k) \quad \text{for } x^{(1)}, \ldots, x^{(k)} \in V
   \]

We can see some classical network models are special cases of this kind of stochastic graph. For example, we have Erdös–Rényi model if $|\Gamma^\Omega| \sim U(0,1)$ and scale-free networks model when $|\Gamma^\Omega|$ follows a power-law distribution. We also have other network models
by using other probability distributions such as Poisson distribution, Geometry distribution, Binomial distribution, etc.

6.2.2. PCM on stochastic graph

Under the PCM on stochastic graph, at each time step \( t \), each node \( x \) generates a random number of neighbours \( \eta \) following a degree distribution given by the Equation \((20)\) and then generates random neighbours set \( GV(x) \) following a point distribution given by the Equation \((21)\); after that \( x \) infects the inactive neighbours \( y \) with a probability \( P_{x,y} \).

The Algorithm 4 presents an example of PCM on Stochastic Graph using power-law distribution for choosing the number of neighbourhoods of each node and discrete uniform distribution for choosing the target nodes.

### Algorithm 4

**Algorithm:** Set \( V = \{1, 2, \ldots, N\} \), stochastic pseudo-closure \( \alpha_0 \), seed set \( A_0 \),

**Require:** powerlaw(a): power-law distribution with parameter a

**Require:** unidrnd(N,a): discrete uniform distribution

1. **procedure** PCM-STOCHASTIC-GRAPH\((V,a,A_0)\)
2. \( t \leftarrow 0, A^{\text{total}} \leftarrow A_0, A^{\text{new}} \leftarrow A_0 \)
3. **while** infection occur do
4. \( t \leftarrow t + 1, A_t \leftarrow \emptyset \)
5. **for** \( u \in A^{\text{new}} \) do
6. \( d_u \leftarrow \text{powerlaw}(a) \quad \triangleright \text{Compute degree of nodes } u \text{ with power-law dis} \)
7. \( N_u \leftarrow \text{unidrnd}(N, d_u) \quad \triangleright \text{Compute } d_u \text{ points randomly from set } V \)
8. \( A_t(u) \leftarrow \{ y^{\text{inactive}} \in N_u, q \leq P(u, y^{\text{inactive}}) \}; q \sim U(0, 1) \)
9. \( A_t \leftarrow A_t \cup A_t(u) \)
10. **end for**
11. \( A^{\text{total}} \leftarrow A^{\text{total}} \cup A_t; A^{\text{new}} \leftarrow A_t \)
12. **end while**
13. **return** \( A^{\text{total}} \) \quad \triangleright \text{Output}
14. **end procedure**

Under the PCM on stochastic graph, at each time step \( t \), each node \( x \in A_{t-1} \) generates a random number of neighbours \( \eta \) following a degree distribution given by the Equation \((20)\) and then generates random neighbours set \( \Gamma_{\ell}(x) \) following a point distribution given by the Equation \((21)\); after that \( x \) infects the inactive neighbours \( y \in \Gamma_{\ell}(x) \) with a probability \( P_{x,y} \).

The Algorithm 4 presents an example of PCM on Stochastic Graph using power-law distribution for choosing the number of neighbourhoods of each node and discrete uniform distribution for choosing the target nodes.

6.2.3. Example

In this section, we illustrate the model presented in the Algorithm 4. In general, the probability density function for power-law\(^2\) is: \( pdf(x, a) = ax^{a-1} \) for \( 0 \leq x \leq \tilde{1}, a > 0 \). The parameter \( a \) is a shape parameter of the distribution. From this probability density function, we will generate a random variable which receive value in \([0, 1]\) from a power law distribution with positive exponent \( a-1 \). In this work, we want to generate the node’s degree distribution which receive value in \([0, x_{\text{max}}]\), \( x_{\text{max}} \) is the maximum value of node’s degree. Therefore, we set up the power-law distribution with the probability density function \( pdf(x, a) = (a/x_{\text{max}}^a)x^{a-1} \) for \( 0 \leq x \leq x_{\text{max}}, a > 0 \). We choose the number of nodes \( N = |V| = 1000 \) and number of initial active nodes \( |A_0| = 10 \). We examine the spreading of the diffusion model with different values of two parameters: \( a \): shape parameter and \( x_{\text{max}} \): the maximum value of node’s degree of the power-law distribution (see Figure 6). For each pair of value of \( a \) and \( x_{\text{max}} \), we run the model 1000 times and compute the average value of active number.

Figure 6(a) represents the result of diffusion process with \( x_{\text{max}} = 10 \) and different values of parameter \( a \) while the Figure 6(b) represents the result of diffusion process with shape parameter \( a=1.5 \) and different values of parameter \( x_{\text{max}} \): the maximum value of node’s degree. We can see in the Figure 6(a) the active numbers in each step with parameter \( a=0.5 \) is significantly lower than other cases. From step 1 to step 5, the greater value of
parameter $a$, the higher value of number of active nodes but from step 6, this trend slightly changes in which the highest value falls in the case $a=1.5$ and maintains the same level until step 15. For the role of the parameter $x_{\text{max}}$, Figure 6(b) shows that the greater value of the maximum of node’s degree $x_{\text{max}}$, the more nodes will be activated in the next step.

\section*{6.3. PCM on multi-relational networks}

\subsection*{6.3.1. Multi-relational network}

A multi-relational network can be represented as a multi-graph, which allows multiple edges between node-pairs. A multi-relational network, which is denoted by $G^{(\text{multi})} = (V, (\Gamma_1, \Gamma_2, \ldots, \Gamma_m))$, is a pair consisting of a set $V$ of vertices and a set of multivalued functions $(\Gamma_i)_{i=1,2,\ldots,m}$ mapping $V$ into $2^V$. $\Gamma_i$ is a neighbour function following the relation $R_i$.

\subsection*{6.3.2. Defining random neighbours set on multi-relational network}

Let us define a random index $\eta$ takes values on $\{1,2,\ldots,m\}$ such as a random variable:

$$P(\eta = i) = p_i; \quad i = 1,2,\ldots,m; \quad p_i \geq 0; \quad \sum_{i=1}^{m} p_i = 1 \quad (22)$$

We define a random neighbour function $\Gamma_\eta$ based on random index $\eta$ such as: $\Gamma_\eta = \Gamma_i$ if $\eta = i$. For each $x \in V$, we can build a random set of neighbours of $x$: $\Gamma_\eta(x) = \Gamma_i(x)$ if $\eta = i, i = 1,2,\ldots,m$.

\subsection*{6.3.3. PCM on multi-relational network}

Under the PCM on multi-relational networks (see Algorithm 5), at each time step $t$, each $x \in A_{t-1}$ generates a random index $\eta$ given by the Equation (22) then generates random neighbours set $\Gamma_\eta(x)$; after that $x$ infects the inactive neighbours $y \in \Gamma_\eta(x)$ with a probability $P_{x,y}$.

We can extend this model by choosing randomly a set $S_\eta \subset \{1,2,\ldots,m\}$ and then using interset or union operator to generate a random set of neighbours of $x$. For example, we

\begin{figure}[h]
\centering
\begin{subfigure}{0.49\textwidth}
\includegraphics[width=\textwidth]{fig1a.png}
\caption{(a)}
\end{subfigure}\hfill
\begin{subfigure}{0.49\textwidth}
\includegraphics[width=\textwidth]{fig1b.png}
\caption{(b)}
\end{subfigure}
\caption{Diffusion process over time for different values of $x_{\text{max}}$ and $a$ with PCM on Stochastic Graph using power-law distribution. (a) Different values of $a \& x_{\text{max}} = 10$. (b) Different values of $x_{\text{max}} \& a=1.5$.}
\end{figure}
can define $\Gamma_\eta(x) = \bigcup_{k \in S_\eta} \Gamma_k(x)$ or $\Gamma_\eta(x) = \bigcap_{k \in S_\eta} \Gamma_k(x)$.

**Algorithm 5** PCM on Multi-Relational Network

Require: $G^{(\text{multi})} = (V, \Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_m])$, seed set $A_0$.
Require: random-index($m, p$): random index distribution, $p = [p_1, p_2, \ldots, p_m]$.

1. procedure PCM-MULTI-RELS-NET($V, p, A_0$)
2.    $t \leftarrow 0$, $A^{\text{total}} \leftarrow A_0$, $A^{\text{new}} \leftarrow A_0$
3.    while infection occur do
4.        $t \leftarrow t + 1$; $A_t \leftarrow \emptyset$
5.        for $u \in A^{\text{new}}$ do
6.            $\eta \leftarrow \text{random - index}(m, p)$ ▷ Compute random index for choosing \nension
7.            $N_u^{\eta} \leftarrow \Gamma_\eta(u)$ ▷ Compute neighbors of $u$ with relation $\eta$
8.            $A_t(u) \leftarrow \{v^{\text{infective}} \in N_u^{\eta}, q \leq P(u, v^{\text{infective}}); q \sim U(0, 1)\}$
9.            $A_t \leftarrow A_t \cup A_t(u)$
10.        end for
11.    $A^{\text{total}} \leftarrow A^{\text{total}} \cup A_t$; $A^{\text{new}} \leftarrow A_t$
12.    end while
13.    return $A^{\text{total}}$ ▷ Output
14. end procedure

**Figure 7.** Example of PCM on multi-relational network: Step 1 and 2. (a) Step 1 - Choosing Neighbours. (b) Step 1 - Diffusion. (c) Step 2 - Choosing Neighbours. (d) Step 2 - Diffusion.
6.3.4. Example
In this section, we illustrate the model on a small multi-relational network built from the Sampson Monastery dataset presented in the Section 3.2. In this dataset, eight relations are coded with four pair of relations which are opposite in their meanings. The relations are esteem and disesteem, like and dislike, positive influence and negative influence, praise and blame.

For diffusion process in this dataset, we consider the semantics of each relation. Therefore, we choose only four relations with positive meanings. We consider in this example the situation in which each person will choose randomly one relation from the positive group of relations and try to infect his neighbourhoods based on the relation he chose. For this reason, in this work, we define multi-relational network with 4 relations: like, esteem, influence, praise and apply PCM-Multi-Rels-Net algorithm to show how information diffusion process takes place in this multi-relational network. We present in the following, the latter steps in more details.

6.3.4.1 Setup. Firstly, we created multi-relational network with 18 nodes represented to the member and nodes connect together with 4 relations. Since, each member ranked only his top three choices on that tie, edges are nonsymmetric, weighted with three integer values, from 1 to 3 and labelled with $R_1, R_2, R_3, R_4$ for like, esteem, influence and praise relation, respectively. We then defined random index distribution $\eta$ such as:

| $\eta$    | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $\sum$ |
|-----------|-------|-------|-------|-------|--------|
| Prob      | 0.35  | 0.3   | 0.2   | 0.15  | 1      |

Since each member ranked only his top three choices on that tie, it is better if we define the probability of diffusion between two nodes based on the weight of the edge instead of...
choosing randomly. Specifically, we assigned propagation probability is 0.4, 0.5, 0.6 if the weight of the edge is 1, 2, 3, respectively.

We begin the diffusion process with a seed set $A_0$. At each time step $t$, each newly active node generates a random neighbours set based on random index distribution and infects the inactive nodes in this set with a probability based on the weight of the edge.

6.3.4.2 Result. Since the diffusion process is calculated probabilistically, we will receive the different result for each time the calculation is made. The Figures 7 and 8 show one of the results of the diffusion process when we run with the seed set $A_0 = [VICTOR_8, AMBROSE_9]$. The two first steps of this example is shown in Figure 7. At the initial time, two nodes VICTOR_8 and AMBROSE_9 are activated. At step 1, node VICTOR_8 and AMBROSE_9 chooses randomly one relation for diffusion by generating random index $\eta$ and has a chance to activate their neighbours following the chosen relation. As we can see in Figure 7(a), VICTOR_8 chooses relation $R_1$ (like relation) and has a chance to active LOUIS_11, MARK_7, ROMUL_10 (AMBROSE_9 is already activated)) while AMBROSE_9 chooses relation $R_3$ (influence relation) has a chance to active BERTH_6 and also LOUIS_11 (VICTOR_8 is also already activated). According to Figure 7(b), only three nodes MARK_7, ROMUL_10, and BERTH_6 are successfully activated and the initial active nodes become old active nodes (denoting it stays active but no chance to activate others). In the next time step, set of newly active nodes [MARK_7, ROMUL_10, BERTH_6] has a chance to active ALBERT_16, AMAND_13, SIMP_18, GREG_2, PETER_4 (Figure 7(c)) and only ALBERT_16, AMAND_13, PETER_4 are successfully activated (Figure 7(d)). The propagation process continues until there is no more new active node. Figure 8 shows the final result of the diffusion process.

7. Conclusion

In this work, we proposed Stochastic Pretopology as a general mathematical framework for complex networks analysis which cannot only deal with uncontrolled factors but also work with a set as a whole entity, not as a combination of elements. We have shown how we construct stochastic pseudo-closure functions to model complex neighbourhoods formation in complex networks in various contexts. Furthermore, we showed how stochastic pretopology can be applied for modelling dynamic processes on complex networks by proposing Pretopology Cascade Model as a general information diffusion model in which complex random neighbourhoods set can be captured by using stochastic pseudo-closure functions. We also illustrated the proposed models by presenting the algorithms and running some experiments. In future works can be developed a software library for implementing stochastic pseudo-closure functions, pretopology cascade diffusion algorithms and applying the proposed models for real-world complex networks.

Notes

1. http://www.casos.cs.cmu.edu/computational_tools/datasets/sets/sampson
2. https://docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.powerlaw.html

Disclosure statement

No potential conflict of interest was reported by the authors.
Notes on contributors

Quang Vu Bui received his Ph.D. in Computer Science, Statistics and Cognition in 2018 from Ecole Pratique des Hautes Etudes, PSL Research University, Paris, France. He is an Assistant professor of Computer Science and Statistics at University of Sciences, Hue University, Vietnam since 2002. His research interests include probability and statistics, complex systems modelling and simulation, topic modeling, complex network analysis, and data sciences.

Soufian Ben Amor received his Ph.D. in Computer Science in 2008 from Ecole Pratique des Hautes Etudes. He is an Associate professor of Computer Science at Versailles University since 2011. He is a member of the team CARO of the PRiSM “Parallelism, Networks and Systems Modeling” Laboratory which is a computer science research laboratory affiliated to both UVSQ (University of Versailles) and CNRS. His research interests include operations research, optimization, complex systems modelling and simulation, and data analysis. Dr. Soufian Ben Amor is now co-leader of the session dedicated to Smart Grids at the French conference on Operations Research ROADEF. He has chaired and/ or served on the advisory boards and program committees of various international conferences and workshops.

Marc Bui is a professor in EPHE, PSL Research University, Paris, France since 2005. His academic work focuses on distributed algorithmics, complex systems modelling and simulation, data sciences and computational humanities. He participated in piloting the ACM-SoICT conference. He co-founded the conference OPODIS, an open forum for the exchange of state-of-the-art knowledge on distributed computing and distributed computer systems.

References

Basileu, C., Amor, S. B., Bui, M., & Lamure, M. (2012). Prétopologie stochastique et réseaux complexes. Studia Informatica Universalis, 10(2), 73–138.
Belmandt, Z. (2011). Basics of pretopology. Paris: Hermann.
Berge, C. (1962). The theory of graphs. London: Courier Corporation.
Bui, Q. V., Amor, S. B., & Bui, M. (2018). Stochastic pretopology as a tool for topological analysis of complex systems. Intelligent Information and Database Systems - 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part II, Lecture Notes in Computer Science, 10752, 102–111.
Bui, Q. V., Sayadi, K., & Bui, M. (2016). A multi-criteria document clustering method based on topic modeling and pseudoclosure function. Informatica, 40(2), 169–180.
Easley, D., & Kleinberg, J. (2010). Networks crowds and markets. Cambridge: Cambridge University Press.
Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 12(3), 211–223.
Granovetter, M. (1978). Threshold models of collective behavior. American Journal of Sociology, 83(6), 1420–1443.
Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. ACM SIGMOD Record, 42(1), 17.
Lamure, M., Bonnevay, S., Bui, M., & Amor, S. B. (2009). A stochastic and pretopological modeling of aerial pollution of an urban area. Studia Informatica Universalis, 7(3), 410–426.
Levorato, V. (2011). Group measures and modeling for social networks. Journal of Complex Systems, 1–10. doi:10.1155/2014/354385
Molchanov, I. (2005). Theory of random sets. London: Springer.
Newman, M. E. J. (2003). The structure and function of complex networks. Siam Review, 45, 167–256.
Newman, M. E. J. (2011). Complex systems: A survey. American Journal of Physics, 79, 800–810.
Nguyen, H. T. (2006). An introduction to random sets. New York: CRC Press.
Sampson, S. (1968). A novitiate in a period of change: An experimental and case study of social relationships. New York: Cornell University.