BEAM: A Network Topology Framework to Detect Weak Signals

Hiba Abou Jamra, Marinette Savonnet, Éric Leclercq
Computer Science Department - EA 7534
University of Burgundy
9, Avenue Alain Savary, F-21078 Dijon - France

***Abstract*—In these days, strategic decision making and immediate action are becoming a complex task for companies and policymakers, since the environment is subject to emerging changes that might include unknown factors. When facing these challenges, companies are exposed to opportunities for growth, but also to threats. Therefore, they seek to explore and analyze large amounts of data to detect emerging changes, or so-called weak signals, that can help maintaining their competitive advantages and shaping up their future operational environments. But due to the increasing volume of daily produced data, scalable and automated computer-aided systems are needed to explore and extract these weak signals. To overcome the automation and scalability challenges, and capture early signs of change in a big data environment, we propose a framework for weak signals detection relying on the network topology. It is implemented under the Cocktail project framework whose goal is to create a real-time observatory of trends, innovations and weak signals circulating in the discourses of the food and health sectors on Twitter. This method analyses quantitatively the network local structure using the graphlets (particular type of motifs) to find weak signals. It provides accordingly qualitative elements that contextualise the identified signals, which will allow business experts to interpret and evaluate their dynamics to determine which ones may have a relevant future. After testing this method on different types of networks (we present two of them in this paper), we proved that it is able to detect weak signals and provides a quantifiable signature that allows better decision making.**

Keywords*—*Weak signals; network analysis; network topology; graphlets

I. Introduction

In recent years, we have noticed an increasing demand to anticipate emerging issues or so-called weak signals, which could help making strategic decisions for future opportunities or threats. In a complex and dynamic world where information circulates rapidly, companies must monitor their environment, in order to adapt and anticipate the market. Monitoring the environment is important in detecting and acting on weak signals, just as it is important to pay attention to transmitted signals at the periphery of the companies' activities without being distracted by the mass of available information. Let's take the example of a soft drink manufacturer who competes with manufacturers of water and new types of beverages instead of traditional colas. The solution to mitigate the risks consists in broadening one's field of vision and thinking by analyzing different sources, such as social networks, blogs and specialized newspapers where one's competitors and customers can express themselves. The information produced by social networks allows decision-makers to have a peripheral and broader vision, and constitutes a source for finding weak signals announcing future threats or opportunities for the company [1], [2]. However, identifying these weak signals usually hidden amidst a large amount of noisy and strong information, and interpreting them when events are not known in advance, becomes a complex problem for researchers [3].

As early as 1975, Ansoff was among the first researchers exploring the field of weak signals. Afterwards in [4], he explained that a weak signal contains partial information available at the time a response is needed. This information must be completed before the signal becomes strong and has an impact on the company. Thus, the identification of a weak signal may announce a future event, its nature (opportunity or threat), its potential impact and the delay before it occurs. The process of carrying out these steps has been described by Ansoff as a graduated response by amplifying weak signals in relation to the level of information. Hiltunen [5] introduced the concept of future sign in which she defined the weak signal using a three-dimensional model: signal, event and interpretation. She aimed to describe a weak signal using the criteria of its visibility, diffusion and amplification. Other terms and definitions qualifying weak signals by author and field of research, are given in Table I.

TABLE I. Different Terms from the Literature to Define a Weak Signal, Classified by Fields of Research

Author	Field of research	Description and terms
Godet (1994)[6]	Economy, strategic prospection	A factor of a big change hardly perceptible at present;
		Component of a strong trend in the future.
Coffman (1997)[7]	Business management studies	Initiator of an important event, of a future trend;
		Signal beyond perception, or within perception but unrecognizable.
Hiltunen (2010)[8]	Event-related future studies	Early information or first symptoms of an emerging event;
		The emerging event itself.
Welz (2012)[3]	Social media, innovation management	Incomplete and imprecise information that is often not perceived;
		Evidence of strong discontinuities of present trajectory.

We hereby choose as definition for a weak signal:

Weak signals are information that provide an indication of upcoming or emerging events that may have significant implications. The information provided is often imprecise, ambiguous and incomplete.

We propose a method called BEAM[1] to detect weak signals and to help business experts in their interpretation. Although this method can be applied on different data sets, we implemented it under the context of Cocktail project to help companies to monitor their image on social networks (especially on Twitter), by detecting weak signals that might be a source of future threats or opportunities. Moreover, we provide a description of weak signal dimensions based on Hiltunen's model. The first two dimensions are translated by the signal support which is a graph built from the data of the study corpus, and a phenomenon/cause[2] which is a precursor of the event. For example, a tweet issued on April 14, 2021 by the official account of Game of Thrones television series, announces a potential remaking of the final season diffused earlier in 2019. The account tweeted "Winter is coming", just as it had done in previous years to announce a new season. Two hours later on the same day, the account again tweeted a video remembering a character's journey in the show. We can consider these tweets as the cause that triggered the weak signals announcing a new (or a remake) season of this show. The third dimension, interpretation, that we use in BEAM, depends on the competencies of experts or decision makers to give meaning to the detected weak signals.

Most weak signal detection approaches study the emergence of keywords using text-mining techniques. Other approaches look to identify the time points at which a behavior of change occurs, so they tend to extract interesting patterns from time series data. Another problem rises is that the data sources used to explore weak signals contain large volumes, which makes their processing and analysis a costly and time-consuming task, so that traditional methods often fail to unveil some of the strategic hidden information. Therefore, it is crucial to adapt a maximum of automation in the detection process. In our proposed approach, we do not consider these techniques neither those that are event detection-based.

Our method relies on the network topology since it allows to explore the local structure of the network and helps finding patterns that can constitute a quantifiable property of the weak signal. Our contributions are the following: 1) we automate the task of weak signals detection, and 2) we provide complementary elements that will help business experts in the interpretation of the detected signals, so that they finally judge their usefulness in a strategic decision or action.

The rest of this paper is organized as follows. Section II presents some of the related works of weak signals detection. We introduce the main steps of our method BEAM in Section III. Section IV presents two use cases and their results and Section V presents some limitations and a discussion. Finally, we conclude the paper and discuss some directions for future work in Section VI.

II. Related Work

In this section, we provide a synthetic description of some major works on weak signal detection, classified according to their detection techniques. Three main categories can be identified: 1) methods based on keyword mapping; 2) methods based on topic modeling; and 3) methods based on network motifs.

The first work is due to Yoon [9] who proposed a quantitative method based on text mining with keyword analysis to identify topics as weak signals in the field of solar panels. The method relies in the search of keywords having a low frequency of occurrence which reflects the visibility of the signal, and a high and increasing rate corresponding to its diffusion. The developed method proposes two metrics: the degree of visibility (DoV) represents the frequency of the keywords in the set of documents, the degree of diffusion (DoD) represents the frequency of the documents. A keyword with a low DoV and a low DoD is considered as a weak signal. He used these metrics to build the Keywords Emergence Map (KEM) and the Keywords Issue Map (KIM). However, Lee and Park [10] then Krigsholm and Riekkinen [11] raise two pitfalls: 1) the problem of uncertainty when the same keyword is at the limit of several quadrants or appears in different quadrants for both maps; 2) the problem of interpretation that occurs when there are several meanings for a given keyword. In addition, keywords related to weak signals are, in general, isolated terms and the absence of relationships and context limits the information in further interpretation. Griol-Barres et al. [12], [13] use a third measure, the degree of transmission (DoT) to assess the importance of keywords. DoT takes into account the documents' h-index where keywords appear. Kim and Lee's approach [14] is based on "document/keywords" matrices. A weak signal is seen as a rare or unusual keyword (outliers) and it is not related to existing topics (cohesion). Analyses are implemented with the Local Outlier Factor (LOF) algorithm [15].

The main interest of using topic modeling techniques is to infer topics from a corpus of documents. In [16], Maitre et al. used Latent Dirichlet Allocation (LDA) with Word2Vec to detect a topic related to a weak signal. For these authors, a weak signal is characterized by a small number of words per document, present in few documents, and unrelated to other topics. Other works have also focused on the detection of weak signals using dynamic LDA as in [17] and [18].

Some researchers have been interested in identifying specific patterns in the networks, called motifs, which could be considered as precursors of events. Baiesi [19] presented a method that studies the correlations between graphs from earthquakes, using tools on network theory. He found that simple patterns such as triangles are an interesting type of major event precursors because they were found in all three studied earthquakes. Kwon et al. [20] use Minimum Spanning Tree (MST) and apply betweenness centrality on the graphs. They consider that a weak signal appears as the node with the lowest centrality score. Moreover, Kim et al. [21] propose the NEST model that collects information from expert networks worldwide. They applied clustering, pattern recognition, and cross-impact analysis using a Bayesian network.

Others considered that weak signals can characterize real-world events, therefore the authors in [22] used tweets frequency and sentiment analysis in an attempt to detect weak signals, using tweets collected during the London riots in 2011. Another study using keyword and sentiment analysis was discussed in [23], where the authors proposed a model

[1]BEAM: a ray or shaft of light beams from the searchlights, a collection of nearly parallel rays (such as X-rays) or a stream of particles (such as electrons), a constant directional radio signal transmitted for the guidance of pilots.

[2]phenomenon = observed fact, normal or surprising event

for predicting a movie performance after its release, using the tweets frequency rates and sentiment analysis.

In contrast to the above studies, our work considers the diffusion and amplification of the signal as important criteria in the detection process (as seen in [5]). In addition, when dealing with social data, it is difficult to apply topic modeling and LDA techniques. Social data consists of short texts (for example 280 characters at most in a tweet), and often contains abbreviations and spelling errors. Moreover, as opposed to these works, our method is not limited to user-generated content analysis but considers a multi-dimensional approach consisting of various data analysis and visualization methods - both quantitative (detection) and qualitative (interpretation). Therefore, we rely on the works related to the identification of specific patterns from the network, since they support our hypothesis that graphlets, which are network particular patterns, can be weak signals.

III. Proposed Framework

Our goal is to provide a weak signal signature; this signature is a quantifiable property of signals that is characteristic of a weak signal [24]. We want to detect weak signals from social network data particularly, that can be represented as an interaction graph between entities. Therefore, the proposed method is based on a topological analysis of the network, in which we have chosen the graphlets as an operational description of weak signals. Indeed, graphlets present characteristics generally associated with weak signals. According to [25] and [26] they are:

- fragmentary and not very visible because they are small sub-graphs;
- meaningless, in fact considering only a sub-graph of at most 5 nodes does not mean much in the mass of data produced by social networks;
- interpretable by business experts by means of their predefined shapes and orbits.

We assume that graphlets allow to automate the task of detecting weak signals in a large volume of data while leaving room for interpretation by experts, thus eliminating the black box effect that a fully automated method could have.

Our method is based on a standard data processing pipeline that includes data collection and exploration, before the detection of weak signals, as illustrated in Fig. 1 and detailed in the rest of this section. In the first step, we prepare the collected raw data, and build the study corpus which is then divided into snapshots of the same duration, for a more refined study. In the second step, we examine the data of the study corpus by applying algorithms and measures for network topology analysis for each snapshot. Then, we identify precursors using different filtering and selection methods. A precursor is an observable and clear fact present in an organization's operational process and caused by existing factors in the process [27]. In the last two steps, we define criteria to qualify the identified precursors as weak signals, and provide contextual elements to interpret them.

A. Raw Data Preparation

BEAM applies to all data types that can be modeled in the form of a graph. For example, we can find such data on the platform https://www.data.gouv.fr/fr/ which makes available French public data. Data from social networks like Twitter and Facebook can be modeled as a labelled graph. However, their semantics is complex because the relations are multiple and context dependent. For example, thanks to a graph we can represent relationships between individuals that can be reciprocal or not, hostile or favorable. Specialized forums and scientific blogs like StackOverflow or MathOverflow can also be modeled as a graph representing interactions between entities.

TABLE II. Example of Interactions Between Entities

u	v	t	i
1	2	1643207148	RT
1	2	1643207149	Q
2	3	1643207149	RT
1	3	1643207150	RT
2	1	1643207150	Q
2	3	1643207150	F
2	4	1643207150	R

BEAM takes as input data a list of interactions where each row represents a four-component tuple of the form (u, v, t, i), where u and v are the identifiers of the entities in contact (e.g., accounts of Twitter users, or two hashtags that are in the same tweet), t determines the time at which the interaction took place. The time can take different formats such as a string or a timestamp epoch[3], etc. The i component represents the type of interaction. For example, Twitter has often been modeled as a graph in which the nodes represent users (or hashtags and urls), and the edges represent different types of interactions that exist between the nodes, like a user retweeting (RT), quoting (Q), replying (R), or following (F) another user, or the co-occurrence of hashtags. As well as replies or mentions on other users' Facebook posts, questions and answers on a scientific forum, etc. Table II shows an example of interactions. If several interactions took place at the same time, we find several rows with the same value on the t component in the interaction list. In this table, the first two rows represent interactions between entities 1 and 2 at two different dates ($t = 1643207148$ and $t = 1643207149$) with two different types of interactions. The last four rows of the list represent four existing interactions between different pairs of entities for the same time ($t = 1643207150$) with different types of interactions.

In the following, we describe the first three phases of the data preparation step.

1) Raw Data Cleansing and Filtering: The global corpus is composed of the raw list of interactions before being processed and filtered. Based on the raw list and the context of the study, it is possible to reduce the data in this corpus by specifying criteria to select only the data that matches the chosen criteria, such as filtering tweets containing hashtags about a topic of interest or some interactions of interest i. Several existing interactions (sometimes having different types) between the same pair of entities at the same time, are grouped together.

[3]Number of seconds elapsed since January 1, 1970 at 00:00; e.g. "26/01/2022 15:25:48" corresponds to timestamp 1643207148.

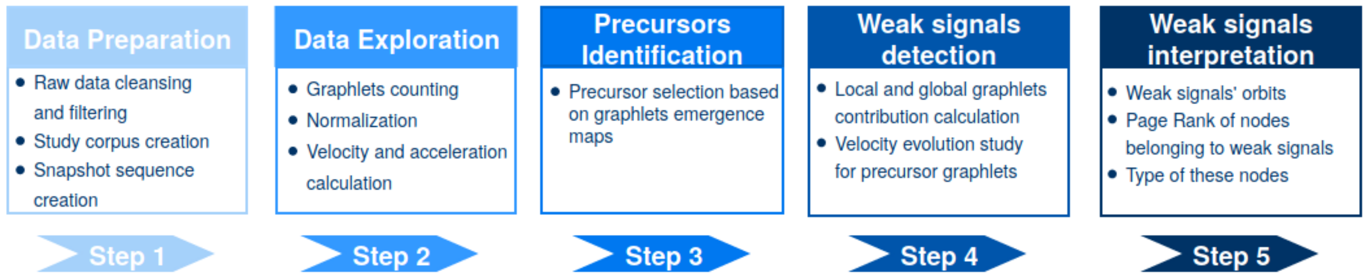


Fig. 1. The Main Steps of our Method BEAM.

This phase is important because the detection of weak signals only makes sense in a well-defined context, and the quality of the filtered data determines the reliability of the decision-making process.

2) Creation of the Study Corpus: Once the data has been filtered according to the chosen criteria, it is transformed into a new set that corresponds to the study corpus. Generally, the study corpus comprises a smaller volume of data than the global corpus. When we do a posteriori weak signals detection, the end of the study period must be determined.

3) Snapshot Sequence Creation: The interaction data can be represented in the form of a graph $G(V, E)$ where V is the set of entities and $E \in V \times V$ is the set of edges modeling the relationship between the pairs of entities, if $(u, v) \in E$ then the nodes u and v are linked together. Temporal interactions between a set of nodes over the time on a period T = $[\alpha, \beta]$ can be formulated by:

$$\mathfrak{X} = \{(u, v, t)\ t \in T \text{ and } u, v \in V\},$$

such that $(u, v, t) \in \mathfrak{X}$ indicates that u interacts with v at time t, these data can be ordered over time t. Temporal interactions do not admit a unique representation: some researchers study them using augmented graphs that integrate temporal information, others still, study them as link streams, others still with a temporal sequence of static graphs [28]. The link stream view was formally defined by Latapy et al. [29]: a link stream $L = (T, V, E)$ is defined by a time interval $T \subset \mathbb{R}$, a set of nodes V, and a set of edges $E \subseteq V \times V \times T$, where $(u, v, t) \in E$, denotes that the nodes u and v have interacted at time t. The augmented graph view represents temporal interactions within a finite set $\mathfrak{X}$, in the form of a single graph. A graph is created in which a node is a pair (v, t), with $t \in T$ and $v \in V$, and in which the node (u, t_i) is related to the node (v, t_j) if they interact, i.e. if $t_i = t_j = t$ and $(u, v, t) \in \mathfrak{X}$, or if they are contiguous in time, i.e. if $u = v$ and $t_j > t_i$ [30]. A second representation is to put temporality on the links.

The representation that we adopted to analyze the temporal interactions is a sequence of $s \in \mathbb{N}$ static graphs: $\mathfrak{S} = \{S^i : i \in \{1, \ldots, s\}\}$ where S^i called snapshot i, is the undirected and unweighted graph (see top part of Fig. 3) containing all the interactions that occurred between times $t_i = \alpha + i\Delta t$ and $t_{i+1} = \alpha + (i+1)\Delta t$, Δt is a constant duration for all snapshots representing one day, one hour, 30 minutes, 10 minutes, etc. The aim of Δt is to connect nodes as a function of time, such that two consecutive snapshots S^t and S^{t+1} are Δt-adjacent. Note that we have grouped the interactions in a time window.

Formally,

$$S^i = (V_i, E_i) \begin{cases} V_i = \{u_i : (u_i, v_i, t) \in \mathfrak{X} \text{ and } t \in [t_i, t_{i+1}[\} \bigcup \\ \qquad \{v_i : (u_i, v_i, t) \in \mathfrak{X} \text{ and } t \in [t_i, t_{i+1}[\} \\ E_i = \{(u_i, v_i) : (u_i, v_i, t) \in \mathfrak{X}, t \in [t_i, t_{i+1}[\\ \text{and } u_i, v_i \in V_i\} \end{cases}$$

Where u_i, v_i are the nodes u and v in the snapshot S^i.

B. Data Exploration

In the following, we explain our BEAM method for finding a quantifiable property of weak signals based on graphlets.

1) Graphlets: Graphlets were first introduced by [31]. A graphlet is a connected induced[4] non-isomorphic[5] subgraph (2 to 5 nodes) chosen among the nodes of a large graph. There are 30 different graphlets ranging from G_0 to G_{29}[6]. An essential element in the context of graphlets are the orbits [25]. They represent the positions (or roles) occupied by the nodes of these subgraphs. There are 73 different positions (from O_0 to O_{72}) for the 30 graphlets. The graphlets up to 4 nodes are presented in Fig. 2, with their corresponding orbits numbered from 0 to 14; in a same graphlet, the orbits having a same color are exchangeable.

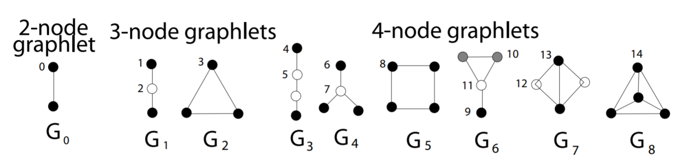


Fig. 2. Representation of the First 9 Graphlets Going from 2 to 4 Nodes.

There are several algorithms for enumerating graphlets and orbits in a graph [32]. In order to choose the most appropriate

[4]In graph theory, an induced subgraph is a subset of the nodes and **all** their connecting edges of an original graph.

[5]In graph theory, an isomorphism of two graphs G and H is a correspondance between the node sets of G and H, such that if any two nodes are adjacent in G, they are adjacent in H. Graphlets are non-isomorphic because they do not have the same form.

[6]In this article, we use the term graphlet for each type out of 30, however it does not represent its occurrence.

algorithm to count graphlets and orbits in the studied graph structures, we defined four criteria:

1) the exact enumeration of graphlets;
2) the enumeration of the orbits;
3) an acceptable complexity;
4) the availability of the source code.

The first criterion ensures the completeness of the graphlets. The second one ensures the interpretability of the results by studying the shape and role of the nodes in the graphlets. Orca satisfies our requirements. It was proposed by Hočevar and Demšar in 2014 [33], which is an exact enumeration algorithm. It counts graphlets up to 5 nodes and enumerates the orbits. It uses an analytical approach based on a matrix representation and works by setting up a system of linear equations per node of the input graph that relates different frequencies of orbits. Its source code is available at: https://rdrr.io/github/alan-turing-institute/network-comparison/src/R/orca_interface.R

With Orca, we count the number of graphlets for each snapshot. Each snapshot S^t is then represented as a vector of thirty elements (G_0^t, G_1^t, G_{29}^t), where G_x^t is the number of the graphlet G_x in the snapshot S^t.

2) Normalization: We apply a normalization procedure on the number of graphlets in order to reduce their values to a particular magnitude. The normalized values are used in the subsequent calculations. This step is of great importance because it should not cover the weak signals but make them comparable to others. The chosen procedure is the one proposed by D. Goldin and P. Kanellakis [34] in which they study the similarity between two queries in a temporal database. A query returns a sequence X of real numbers $(x_1, \ldots, x_n)$. Two reals a, b define a transformation $T_{a,b}$ on X by relating each x_i with $a \times x_i + b$. $\overline{X}$ represents the normal form of X, calculated by:

$$\overline{X} = T^{-1}_{\sigma,\mu}(X) = T_{\frac{1}{\sigma},-\frac{\mu}{\sigma}}(X)$$

Where $\mu(\overline{X})$ = 0 and $\sigma(\overline{X})$ = 1, μ being the mean and σ the standard deviation.

Applying this normalization procedure for each snapshot S^t where s is the number of snapshots, each component G_x^t of its vector with $x \in \{0, \ldots, 29\}$ is therefore normalized by:

$$\overline{G_x^t} = \frac{G_x^t - \mu(G_x)}{\sigma(G_x)}$$

with $\mu(G_x)$ the mean of each graphlet G_x for all snapshots, given by:

$$\mu(G_x) = \frac{1}{s}\sum_{t=1}^{s} G_x^t$$

and $\sigma(G_x)$, the standard deviation, calculated by:

$$\sigma(G_x) = \sqrt{\frac{\sum_{t=1}^{s}(G_x^t - \mu(G_x))^2}{s-1}}$$

3) Estimation of the Signal Reinforcement: Diffusion and Amplification: Velocity and acceleration evolution are quantitative features that allow us to evaluate the diffusion and amplification of the signal. We use these measures to identify event precursors among the graphlets.

From the obtained normalized values, the evolution of all $\overline{G_x^t}$ components is studied via their velocity and their acceleration. Our objective is to highlight the graphlets which emerge before others. For each snapshot and each graphlet G_x, we compute its velocity as the difference between the normalized value of the graphlet in snapshot S^{t+1} and the normalized value of the same graphlet at snapshot S^t. As for the acceleration, it is also calculated in the same way by making the difference between the velocities.

We therefore acquire a numerical matrix representing for each snapshot, the normalized value of the graphlets $\overline{G_x^t}$, their velocity $\overline{V_x^t}$, and their acceleration $\overline{A_x^t}$. This matrix is presented in the Table III which illustrates an example of the values of some graphlets per snapshot. Note that the velocities can only be computed from snapshot S^2, and the accelerations from snapshot S^3.

TABLE III. EXAMPLE OF NORMALIZED VALUES OF GRAPHLETS, THEIR VELOCITY AND ACCELERATION FOR THE SNAPSHOTS OF THE STUDY PERIOD

Snapshot	Graphlet	$\overline{G_x^t}$	$\overline{V_x^t}$	$\overline{A_x^t}$
S^1	G_1	-0.952	NA	NA
S^1	G_2	-0.222	NA	NA
S^2	G_1	-0.666	-0.121	NA
S^2	G_2	0.456	0.678	NA
S^3	G_1	-0.786	0.938	1.059
S^3	G_2	0.758	0.302	-0.376
S^4	G_1	0.152	0.975	0.037
S^4	G_2	0.758	-0.315	-0.617
....	....	....	....	....
S^s	G_{29}	0.503	0.129	0.225

Fig. 3 summarizes the results obtained from the topological characteristics of the graphs for precursor discovery.

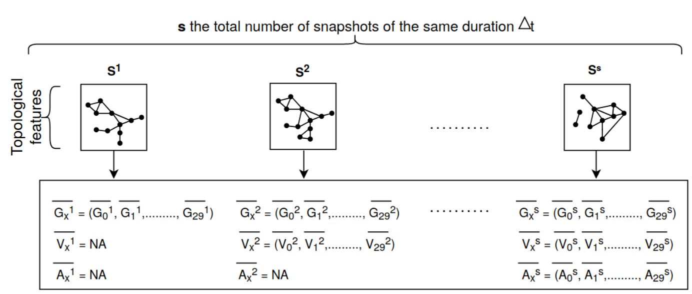


Fig. 3. Representation of the Calculation of Graphlets, Velocities and Accelerations for the Snapshots of the Study.

C. Precursors Identification

First, we present the principles of precursor identification and then its implementation set up. To identify precursors, we rely on Hiltunen's approach in which she considers two types of precursors: early information and first symptoms. The first type represents new information that appears suddenly, such as the announcement of a new product or invention. The second type represents a remarkable change that is difficult to interpret.

We have illustrated this representation in Fig. 4 where we consider four zones/quadrants. The ”noise” area corresponds to graphlets with very small or negligible velocities and accelerations. The three remaining areas correspond to graphlets with 1) high velocity and low acceleration i.e. Q4, 2) low velocity and high acceleration i.e. Q1, 3) high values for both criteria i.e. Q2. If a graphlet is located in the areas Q1 and Q4 of the figure, we consider it as early information, whereas if it is located in the Q2 area, we consider it as first symptoms because the values of its velocity and acceleration are much higher. The graphlets located in these three zones are precursors of events because they are observable and clear facts or remarkable by their velocities and/or accelerations. The graphlets located in the noise zone are not considered as event precursors, because they represent unclear information which appears randomly but is not meaningful.

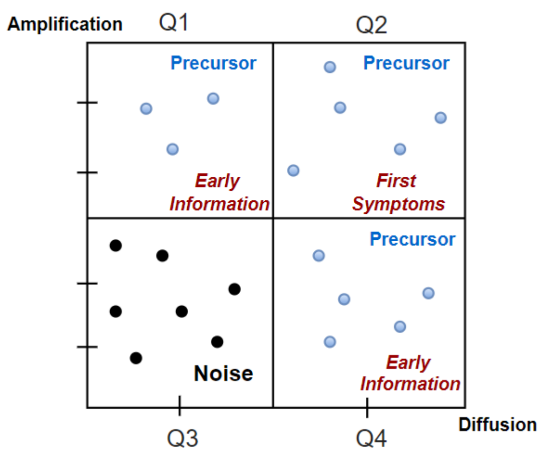


Fig. 4. Structure of the Graphlets’ Emergence Map for a Particular Snapshot S^t.

Our aim here is to propose a division into graphlet emergence zones in the form of the four quadrants, not necessarily of equal size, as presented above. We propose two solutions:

1) **Slicing with respect to a top k:** The first possibility consists for each of the s snapshots S^t in
   - ranking their velocities $\overline{V_0^t}, \overline{V_1^t}, \overline{V_2^t}, \dots, \overline{V_{29}^t}$ in ascending order;
   - settting a top k value for which the k graphlets have the highest velocities.

   Similarly, we set a top k for the graphlet acceleration criterion.
2) **Slicing with respect to a threshold μ:** The second possibility consists in setting a threshold for each of the s snapshots S^t equal for example to the average of the velocities and in selecting the velocities greater than or equal to this value, $\overline{V^t} \geq \mu(\overline{V^t})$. A similar division is made for the accelerations.

For a snapshot S^t, we create an emergence map that represents on the x-axis the velocity values, and on the y-axis the acceleration values. The structure of this map is based on the representation of the signals as seen in Fig. 4, where we place the graphlets in the four quadrants of the map. The division into quadrants depends on the choice of one of the two slicing methods presented above.

D. Identification of Weak Signals

The aim of this step is to qualify the graphlets as weak signals. We want to keep the True Positives and introduce the True Negatives as weak signals. False Negatives are graphlets that are neither precursors nor weak signals, and False Positives are graphlets that have been identified as precursors, but are not weak signals, they should be eliminated at the end of this step. The Table IV qualifies the results obtained by BEAM and estimates them.

TABLE IV. Qualification of the Results Obtained by BEAM

	Weak signal	¬ **Weak signal** (noise, strong signal)
Precursor	True positive	False positive
¬ **Precursor**	True negative	False negative

To qualify graphlets as weak signals, our BEAM method relies on:

1) Graphlets contribution in weak signals;
2) Evolution of the signal’s diffusion.

The first measure aims to detect the weak signals at a certain time t, and the second one confirms or denies the detected signals, for an interval of time following t.

1) Contribution Ratio Calculation: We look in this phase to measure the contribution of a graphlet according to all graphlets to verify whether they are weak signals or not. The qualification criterion here is a ratio calculation. We propose two calculations: a local ratio that takes into account only the current time window, and a global ratio that takes into account all the studied windows. The number of graphlets in the chosen time window t is divided by the total number of graphlets for this time window:

$$R_{Local}(G_x) = \frac{G_x^t}{T_{Local}(G)}$$

Where $T_{Local}(G) = \sum_{x=0}^{29} G_x^t$.

The total number of a graphlet in the set of studied snapshots is divided by the total number of graphlets for all this set of snapshots, as follows:

$$R_{Global}(G_x) = \frac{\sum_{t=1}^{s} G_x^t}{T_{Global}(G)}$$

Where $T_{Global}(G) = \sum_{t=1}^{s}(\sum_{x=0}^{29} G_x^t)$, with s the number of processed time windows.

The resulting ratios $R_{Local}(G_x)$, $R_{Global}(G_x)$ are ranked in ascending order to qualify graphlets as weak signals if they are at the top of the list; the other graphlets are eliminated. This calculation is related to the **rareness characteristic** of weak signals since it selects the weakest ratios. Finally, at the end of these different analyses, we can provide per snapshot S^t a signature of the weak signal as a vector of at most 29 components: $(G_x^t, \dots, G_y^t)$ with $x, y \in \{0, \dots, 29\}$.

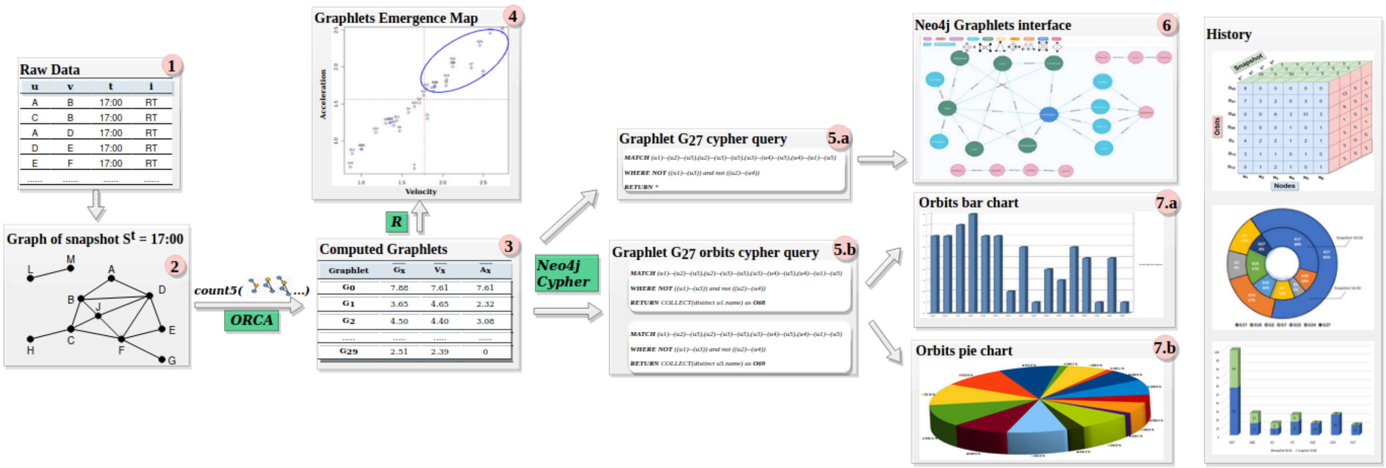


Fig. 5. BEAM user Interface for a Particular Snapshot. On the Right a Historical Panel for All Snapshots.

2) Evolution of Graphlet Velocities: The aim here is to propose a visual support to the identification of weak signals that can be carried out in parallel with the creation of emergence maps. This support consists in monitoring the evolution of the graphlet velocities over the time. To do this, we build a heatmap to visualize the evolution of the values of this criterion over a window of snapshots surrounding S^t : $\left[S^{t-1}, S^{t+3}\right]$. For example, if a graphlet G_x is found among the precursors in S^t, and if its velocity decreases in S^{t+1} and this decrease remains in S^{t+2}, then we consider that it is a false alarm thus we will not qualify it as a weak signal. On the contrary, the graphlets whose velocity increases in the studied interval, are True Positives and qualified as weak signals. Similarly, those that were not identified among the precursors but their speed increases, are True Negatives.

E. Interpretation of Weak Signals

Searching for relevant data for important event anticipation has several challenges including the complexity of detection and interpretation. At first sight, weak signals are fragmented pieces of information, so hard to be interpreted. However, if they were combined with context or additional elements, the interpretation task may become clearer and more relevant. Hiltunen [5] considers that the interpretation dimension means the information user's understanding of the future signal.

Our aim in this step is to make sense of the identified weak signals to help business experts in better and useful decision making. Several works tried to address this dimension like Lee and Park [10] who employed keyword clustering and topic selection based on keyword co-occurrences, or J. Kim [14] who proposed to observe the rarity and the paradigm unrelatedness of weak signals. In BEAM, we consider elements characterizing the nodes like their name, their type, as well as their role, etc. The role of a node can be characterized by its orbit, or its position in the graphlets. We therefore take advantage of this criterion to identify the central, cross or internal nodes roles. With Orca, we can also count how many times a node is playing these roles. BEAM also provides to business experts different visualizations that may help interpreting the identified signals, as shown in Fig. 5. This figure illustrates an example of one snapshot being studied by an expert, where parts 1 and 2 represent the raw interactions data at this snapshot, and the corresponding generated graph. Once Orca is executed, it allows the user to fetch all computed graphlets (their normalized number, their velocity and their acceleration) shown in part 3, and allows him to spot the precursor graphlets on the emergence map (part 4). In parts 5.a and 5.b, the expert may look further for a particular precursor graphlet using the Neo4j Cypher queries implemented in BEAM. Consequently, he has two visualization options, either 1) with a view that highlights the precursor graphlet (here G_{27} for example) in the initial graph, as in part 6, or 2) with a chart view representing the number of nodes in each of the precursors' orbits (including G_{27}), as shown in the bar and pie charts (parts 7.a and 7.b). At the right of this figure, there is a historical panel that allows the user to go back to a visited visualization, or to visualize the data based on a sequence of snapshots (in the form of a data cube, a clustered bar, or pie chart).

IV. Empirical Case Study

In order to evaluate the proposed method, we performed experiments on different datasets. We will describe in this section the empirical studies carried out on the Game of Thrones Twitter network (GOT), and a sensor network holding interactions between elementary school students. The purpose behind these studies is to 1) ensure the robustness of the method by verifying that weak signals are detected prior to a certain event (GOT), but are not detected when the event is already foreseen and punctual (elementary school); and 2) confirm its reproducibility through 6 episodes of GOT.

A. Game of Thrones Dataset

With long-tail streaming, HBO has pegged the last season of GOT as averaging around 44.2 million viewers per episode. The final season of GOT has upset many viewers because of changes in writers, shortening the season (this last one had only six episodes diffused once per week), and a surprising turn of events. This is why the fan club was more into criticizing this final season. It was released on April 14 2019, and ended on May 19 2019. The episodes were diffused live at 9 p.m.

American time, and the collected data corresponds to the tweets published at the same timestamp. The original data sample including all episodes of the season, has a total number of 46 481 705 tweets including 34 094 365 retweets.

1) Study Corpus: We describe the study carried on the first episode with a period limited to 14-04-2019 from 1 p.m. to 7 p.m. The graph represents 223 383 retweets between 189 741 users corresponding to audience, journalists, media, etc. This graph is divided into a sequence of snapshots with a duration Δt equal to 10 minutes.

2) Precursors and Weak Signals: For each snapshot, we counted the number of 5-node graphlets and normalized them, then computed their corresponding velocity and acceleration. In the following, we will present only the obtained results for the snapshot of 5 p.m. Fig. 6 illustrates the graphlets'

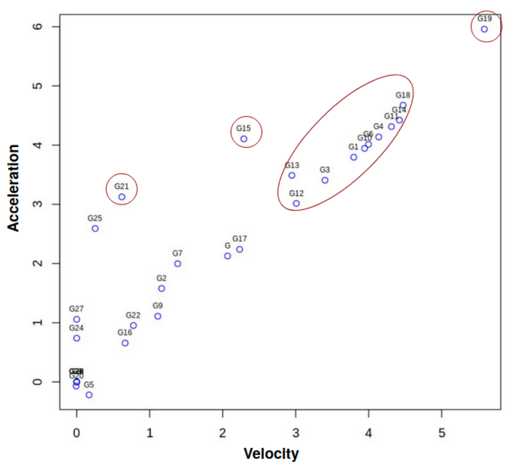


Fig. 6. Precursors Selected Among the Graphlets in the Snapshot of 5 p.m.

emergence map of 5 p.m built based on these criteria. The 13 graphlets that are circled in red are selected as precursors, since they belong to one of the aforementioned zones Q1, Q2 or Q4. The remaining graphlets are considered noise being positioned in the Q3 zone.

Once the precursors are selected, the next step is to check their relevance by identifying the false positives, preserving the true positives and adding the true negatives. For this, we first calculated the local and the global contributions of the graphlets for each snapshot, and ranked the resulted ratios in ascending order. For the 5 p.m. snapshot, we selected the top 5 graphlets according to the ranked ratios, that are respectively G_{21}, G_{15}, G_{13}, G_{12} and G_3. We moved next to confirm the obtained contribution ratios, by analyzing the evolution of graphlets' velocities in the snapshots surrounding 5 p.m.

We monitored the evolution of graphlets (precursor ones in particular) velocities during an interval of snapshots going from 4:50 to 5:30 p.m., as shows the Fig. 7. The x-axis of the heatmap in this figure represents the different snapshots, and the y-axis represents some of the precursor graphlets. The legend on the right of the figure represents the graphlets velocity values between -2 and +5. The blue colors in the heatmap are the graphlets G_x such that $\overline{V_x^t} \geq +5$, and the red colors are those such that $\overline{V_x^t} < 0$. For example, when considering the 5:00 p.m. snapshot, the graphlet G_{19} shows

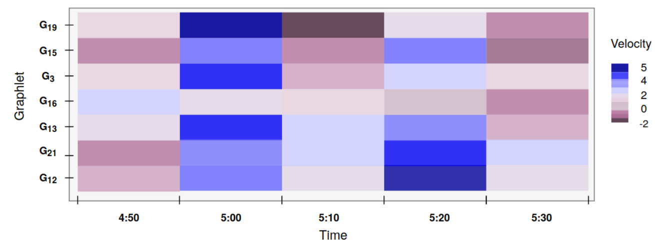
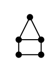
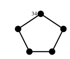
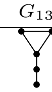
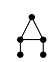


Fig. 7. Velocity Evolution of Some Precursors in the Snapshots Around 5 p.m.

a significant increase in velocity up to +5, and in the next snapshot, it decreases rapidly to -2. Another evolution is the one of G_{16}, that shows continuous decrease in its velocity for all snapshots. We revised the global contributions of these graphlets and noticed that they have higher ratios than the top 5 mentioned above. These results are therefore consistent, and indicate that these graphlets are false alarms. As for the remaining graphlets, we see that they show a significant rise in their velocities at 5 p.m., followed by a minor decrease at 5:10 p.m., then another rise at 5:20 p.m. The decreases at 5:30 p.m. do not affect the analysis results, because it is enough to observe one remarkable rise after the current snapshot (i.e. 5 p.m.). These last results are consistent with those obtained from the contribution calculation, hence we qualify the top 5 graphlets from the list as weak signals. Table V illustrates the five qualified weak signals, their shape and their global contribution.

TABLE V. TOP 5 GRAPHLETS QUALIFIED AS WEAK SIGNALS FOR THE SNAPSHOT OF 5 P.M.

Graphlet	G_{21}	G_{15}	G_{13}	G_{12}	G_3
Graphlet shape					
Graphlet global contribution	0.0009	0.0174	0.0235	0.0260	0.0283

3) Weak Signals Interpretation: Our objective here is to help in the interpretation of identified weak signals, by providing contextual elements like user accounts. We first start by calculating the number of orbits (nodes positions) of graphlets with Orca, then analyze these positions in weak signal graphlets. This is an important task because it can help explaining the users' role in assessing and supporting this final season.

Table VI illustrates an extract of the obtained results from this step at the 5 p.m. snapshot. The accounts `@dcucomics`, `@jonatas_maia12` and `@9GAG` appear mostly in a peripheral position in graphlet G_3 (O_4). The remaining accounts in this table occupy a central position (O_5). The third and fourth columns of the table represent respectively the different node types (from comic accounts, to online platforms or journalists), and their corresponding ranks after executing the PageRank algorithm on 7000 users. We note that the official accounts of the series are communicating more than the others since they occupy the most central positions. In addition, the account `@TylerIAm` has a you-tube channel and diffuses live broadcasts from 12 to 3 p.m., so from his PageRank we can see that he retweeted very much about GOT two hours after the broadcast.

TABLE VI. Extract of Important Nodes Belonging to G_3 Orbits, and their Page Rank for the Snapshot of 5 p.m.

Graphlet-Orbit	Account	Account type	PageRank
O4	@dcucomics	DC universe (fictional universe produced by Warner Bros) fan account	27
	@jonatas_maia12	Product designer	61
	@9GAG	Online platform, viral and funny videos	248
O5	@GameofThrones	Official account of GOT	1
	@Thrones_Memes	GOT memes account	3
	@TylerIAm	Journalist	4
	@LordSnow	Got character (John Snow)	12

We then conducted a refined analysis on the account `@TylerIAm` using the Neo4j Cypher queries implemented in BEAM. Here we went from a graphlet type to 4 graphlet instances, and displayed the users linked to this account on a G_3 instance. As shows Fig. 8, `@Woodlawnwonder`, a blogger, occupies the other central position of G_3, and the peripheral positions are occupied by the remaining accounts. We considered the instance containing `@joestudz18`, an artist and concept illustrator, as well as `@holy_schnitt` who is a social media star. We assume that the blogger account activity for example, should be monitored by business experts.

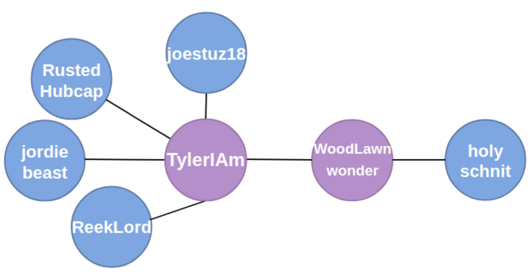


Fig. 8. Important Accounts Occupying Orbits of an Instance of G_3 in the Snapshot of 5 p.m.

We also noticed an interesting fact about nodes appearing in G_{15}. These nodes belong to accounts located in Brazil, including a comic and digital creator `@cleytu`, an entertainer `@fabwiano`, the Brazilian version of HBO channel `@HBO_Brasil`, and a journalist `@LethyciaDias_`, etc. We then executed the Louvain algorithm for the whole episode (where all snapshots are combined). We noticed that all these users belong to the same community. We therefore allow business experts to give their final assessment regarding these results and the type/location of the users, since the shape of this graphlet can put into evidence an important social structure [26].

We performed the same experiment on the retweets published on the remaining 5 episodes of GOT, to test the reproducibility of our method. The results obtained showed that the weak signal graphlets are detected few hours before the episode is broadcast. Moreover, we noticed that the same graphlets are found in several episodes as weak signals.

According to the results of our study, the official GOT accounts (`@GameofThrones`, `@ThronesMemes` and `@LordSnow`) communicated supporting information (they were motivating the audience to watch GOT, and highlighting the crucial role of the character Jon Snow). Therefore, the weak signals carried by these accounts can be opportunities, which is important that they emerge to verify if the speech is well perceived. On the other hand the comic accounts `@dcucomics` and `@9GAG` can be threats because they can make fun of GOT or even divert its image, so they must be watched to mitigate their negative impact on the audience.

B. Elementary School Interactions Dataset

We conducted another experiment on students interactions in an elementary school in Lyon, France, on two consecutive days in October 2009 [35]. The school day runs from 8:30 a.m. to 5:30 p.m., with a lunch break between 12:00 p.m. and 2:00 p.m., and two 20-25 minute breaks at around 10:30 a.m. and 3:30 p.m. that take place in a common playground. Two students are considered to be interacting if they are within 3 to 5 feet of each other for at least 20 seconds.

Through this experiment, we aimed to:

1) verify that no weak signal is found when the events are known (as the breaks are scheduled at specific times, without particular triggers, . . .);
2) verify that the error rate is reduced, i.e. that the false positives are eliminated by our method.

1) Study Corpus: The raw data file was downloaded from the official site[7]. We divided the corpus into snapshots of 10 minutes duration each, and choose to work with five snapshots that correspond to the times before the lunch break (from 11:20 a.m. to 12 p.m.) of the first school day.

2) Precursors and Weak Signals: We focused our analysis on the snapshot of 11:40 a.m. and we were able to spot precursors on the corresponding graphlets' emergence map. Thereafter, we evaluated the located precursors using the graphlets contribution ratios, as well as their velocity evolution. The velocities heatmap showed that for this particular snapshot, the graphlets reach high values, but then decrease rapidly in the next snapshot (11:50 a.m.) and continue to decay until the last snapshot. This decrease supports the hypothesis that the identified precursors at this snapshot were false alarms, hence shall not be qualified to weak signals. Although we did not expose weak signals from this experiment, the dataset is a ground truth example on which we relied to confirm the method's objectives.

C. Results Validation: Cross-Correlation

This is a supplementary step that aims to validate the discovery of weak signals by studying the relation between them and a potential event[8]. In this step, we observe the data no more as a sequence of graphs, but instead as time series.

[7] http://www.sociopatterns.org/datasets/primary-school-temporal-network-data/
[8] We consider an event as a situation in which the number of interactions reaches its maximum value.

A time series X is created from the number of interactions selected between all pairs of nodes. It is a sequence of n elements $X = (x_i)_{1 \leq i \leq n} = (x_1, x_2, \ldots, x_n)$.

We rely on the Cross-correlation to validate the intrinsic properties of BEAM by studying the relation between the graphlets weak signals and the original time series of interactions (built from the raw data). Cross-correlation[9] is a linear measure of similarities between two time series X and Y, which helps evaluate the relation between these series over time [36]. An offset/lag h is associated with this measure, knowing that if $h < 0$ then X could predict Y, and if $h > 0$ then Y could predict X.

We applied this function on the GOT dataset, by first building the time series corresponding to the users' retweets, then building a time series for each one of the 30 graphlets. From the obtained results, we noticed that the graphlets that were qualified as weak signals (see Table V), presented positive correlations (between 0.5 and 0.7) with the initial series with a lag of 10 or 20 minutes (Δt = 10 minutes). Fig. 9 illustrates the correlogram of graphlet G_{13} with the initial retweets series. The graphlets identified as false alarms in the previous steps of the method, like G_{16} and G_{19} for instance, did not present correlations.

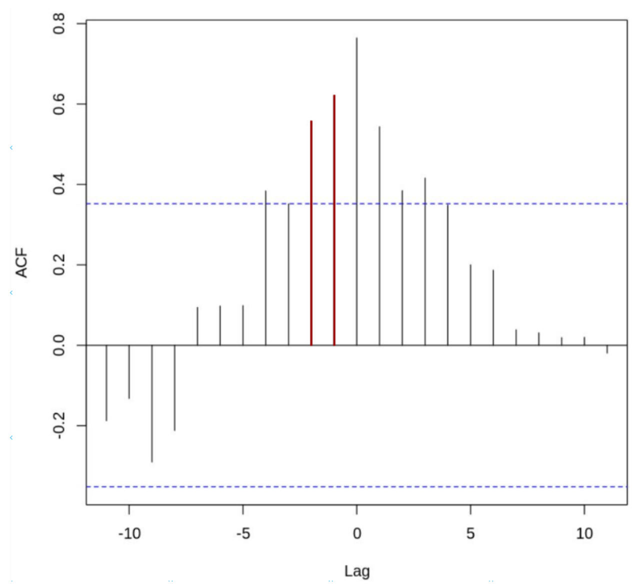


Fig. 9. Positive Correlation with Negative Lags of 10 and 20 Min Between G_{13} and the Retweets Time Series (GOT Dataset.)

We executed the Cross-correlation on the elementary students interactions dataset, and from the obtained correlograms, we did not find any negative-lag correlation between the initial time series and the precursor graphlets. These results confirm that when the event is already planned and punctual, there are no weak signals that can announce the occurrence of such event, hence the detection phenomena corresponds in this case to false positives that we were able to identify.

[9] implemented with the R package *tseries*: https://www.rdocumentation.org/packages/tseries/versions/0.1-2/topics/ccf

V. Limitations and Discussion

Several works relied on different techniques to detect future weak signals. After applying BEAM on several data sets, the findings have been quite consistent in establishing that this method is able to detect weak signals prior to an important event. Using BEAM, the business expert does not spend time anymore to choose selection filters, prepare data and extract remarkable information for analysis. Instead, he relies on the previously detected information and assesses if it should be qualified as future weak signals or not.

However, BEAM still presents some limitations related to: 1) The constitution of study corpus. 2) The snapshots duration Δt. 3) The division of graphlets' emergence maps. 4) The interpretation of the detected signal through its recognition.

These limitations are related to filters that hinder the analysis of weak signals and constitute barriers to their interpretation. These barriers were discussed by Ansoff [4] who suggested that weak signals must pass three filters 1) the monitoring (surveillance), 2) the mentality and 3) the power filter, before potentially triggering an action or a decision. The monitoring filter corresponds to the capacity of the weak signal to be detected or discovered, in the midst of all other perceived information, by one or more actors within the organization. The mentality filter refers to the capacity of the signal to be recognized, after being detected, as relevant information with regard to the examined situation. Finally, the power filter refers to decision-making once the signal has been detected and its relevance recognized. The people in charge in the organization can decide for example not to make this signal a priority, despite the underlying risk.

Limits 1, 2 and 3 of BEAM are related to the monitoring filter. The first one results from an irrelevant choice of keywords, accounts or hashtags (in case of Twitter data) during the creation of the study corpus. It can be resolved by adding a feedback loop in BEAM, to allow business experts to return to the first step and modify the selection criteria, if the interpreted signals are not significant for them. Regarding the second one, the duration of snapshots Δt should be well-chosen for an expert to analyze the detected signals and provide his decision. To do so, we performed an experiment to measure the time in which a tweet becomes viral (i.e. when it reaches the highest number of its retweets). This limitation is also depending on the number of tweets in the studied snapshots, if it is high, Δt should be limited otherwise increased. The third limitation relates to the choice of the top k or the threshold that divides the graphlets' emergence maps for the precursors selection. This one was resolved with the contribution calculation with which graphlets are added or removed from the list of weak signals based on their ratio. The fourth limitation is related to the mentality and power filters, thus cannot be resolved within BEAM, since it is up to the business expert to provide his final interpretation and decision regarding the detected signals.

VI. Conclusion

In this article, we presented our method BEAM whose objective is to help business experts to detect and interpret the identified weak signals in order to enable them to make decisions and plan future strategies. First, we find graphlets in a graph of interactions between entities, which represent clear and observable facts, quantifiable using measures of diffusion and amplification that characterize them as precursors. We examine the contribution of graphlets to eliminate false alarms (False Positives) and qualify True Positives and True Negatives as weak signals. Once these signals are identified, the shapes of the graphlets and their orbits help the business experts in their interpretation. In BEAM, we have chosen graphlets as an operational description to give a signature of the weak signal, this choice allows an automation of the detection task while enabling a final judgment by the business experts for a better decision making.

Next, new experiments will be performed on other types and larger networks to resolve the remaining limitations of the method. In further research, we would like to explore if graphlets can be used to indicate phase transitions of an information emergence between transition points. This will help in analyzing the weak signal amplification process. BEAM combines data analysis and visualization tools to guide business experts in detecting and interpreting weak signals, and offers a great potential for decision-making in most business organizations. Moreover, the detection of weak signals offers promising leads for innovation.

Acknowledgment

This work is supported by the Investissements d'Avenir program, Cocktail ISITE-BFC project (Initiatives Science Innovation Territoires Économie en Bourgogne-Franche-Comté), ANR contract 15-IDEX-0003, https://projet-cocktail.fr/.

References

[1] G. Day and P. Schoemaker, "Peripheral Vision: Sensing and Acting on Weak Signals," *Long Range Planning*, vol. 37, p. 117–121, 04 2004.

[2] J. Wu and A. Pinsonneault, "Enhancing peripheral vision through social media use: A social network perspective," in *AMCIS*, 2011.

[3] K. Welz, L. Brecht, A. Pengl, J. V. Kauffeldt, and D. R. Schallmo, "Weak signals detection: Criteria for social media monitoring tools," in *ISPIM Innovation Symposium*. The International Society for Professional Innovation Management (ISPIM), 2012, p. 1.

[4] H. I. Ansoff and E. J. McDonnell, "Implanting strategic management," 1990.

[5] E. Hiltunen, "The future sign and its three dimensions," *Futures*, vol. 40, no. 3, pp. 247–260, 2008.

[6] M. Godet, *From anticipation to action: a handbook of strategic prospective*. UNESCO publishing, 1994.

[7] B. Coffman, "Weak signal research, part I: Introduction," *Journal of Transition Management*, vol. 2, no. 1, 1997.

[8] E. Hiltunen, "Weak Signals in Organisational Futures," *Aalto University School of Economics, Aalto*, 2010.

[9] J. Yoon, "Detecting weak signals for long-term business opportunities using text mining of Web news," *Expert Systems with Applications*.

[10] Y.-J. Lee and J.-Y. Park, "Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence," *Quality & Quantity*, vol. 52, no. 2, pp. 653–667, 2018.

[11] P. Krigsholm and K. Riekkinen, "Applying Text Mining for Identifying Future Signals of Land Administration," *Land*, vol. 8, no. 12, 2019. [Online]. Available: https://www.mdpi.com/2073-445X/8/12/181

[12] I. Griol-Barres, S. Milla, and J. Millet-Roig, "System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques," in *ICAART*, 2019.

[13] I. Griol-Barres, S. Milla, A. Cebrián, H. Fan, and J. Millet, "Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing," *Sustainability*, vol. 12, no. 19, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/19/7848

[14] J. Kim and C. Lee, "Novelty-focused weak signal detection in futuristic data: Assessing the rarity and paradigm unrelatedness of signals," *Technological Forecasting and Social Change*, vol. 120, pp. 59–76, 2017.

[15] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 93–104.

[16] J. Maitre, M. Ménard, G. Chiron, A. Bouju, and N. Sidère, "A Meaningful Information Extraction System for Interactive Analysis of Documents," in *International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 92–99.

[17] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 113–120. [Online]. Available: https://doi.org/10.1145/1143844.1143859

[18] L. Pépin, P. Kuntz, J. Blanchard, F. Guillet, and P. Suignard, "Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets," *Computers & Industrial Engineering*, vol. 112, pp. 450–458, 2017. [Online]. Available: https://doi.org/10.1016/j.cie.2017.01.025

[19] M. Baiesi, "Scaling and precursor motifs in earthquake networks," *Physica A: statistical mechanics and its applications*, vol. 360, no. 2, pp. 534–542, 2006.

[20] L.-N. Kwon, J.-H. Park, Y.-H. Moon, B. Lee, Y. Shin, and Y.-K. Kim, "Weak signal detecting of industry convergence using information of products and services of global listed companies-focusing on growth engine industry in south korea," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 4, no. 1, p. 10, 2018.

[21] S. Kim, Y. Kwon, Y. Jeong, S.-B. Choi, J.-K. Park, and S.-W. Hong, "Nest: a model for detecting weak signals of emerging trends using global monitoring expert network." in *EKAW (Posters and Demos)*. Citeseer, 2010.

[22] C. Charitonidis, A. Rashid, and P. J. Taylor, "Weak signals as predictors of real-world phenomena in social media," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 2015, pp. 864–871.

[23] S. Asur and B. A. Huberman, "Predicting the future with social media," in *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology*, vol. 1. IEEE, 2010, pp. 492–499.

[24] D. Georgiadis and M. Raubal, "An interdisciplinary review on weak signal detection: Future resilient systems working paper# 3," 2020.

[25] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, pp. e177–e183, 2007.

[26] R. Charbey, L. Brisson, C. Bothorel, P. Ruffieux, S. Garlatti, J.-M. Gilliot, and A. Mallégol, "Roles in social interactions: graphlets in temporal networks applied to learning analytics," in *International Conference on Complex Networks and Their Applications*. Springer, 2019, pp. 507–518.

[27] P. Korvers, "Accident precursors: Pro-active identification of safety risks in the chemical process industry." 2004.

[28] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, p. 97–125, Oct 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2012.03.001

[29] M. Latapy, T. Viard, and C. Magnien, "Stream Graphs and Link Streams for the Modeling of Interactions over Time," *CoRR*, vol. abs/1710.04073, 2017. [Online]. Available: http://arxiv.org/abs/1710.04073

[30] V. Kostakos, "Temporal graphs," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 6, p. 1007–1023, Mar 2009. [Online]. Available: http://dx.doi.org/10.1016/j.physa.2008.11.021

[31] N. Pržulj, D. G. Corneil, and I. Jurisica, "Modeling interactome: scale-free or geometric?" *Bioinformatics*, vol. 20, no. 18, pp. 3508–3515, 2004.

[32] P. Ribeiro, P. Paredes, M. E. Silva, D. Aparicio, and F. Silva, "A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–36, 2021.

[33] T. Hočevar and J. Demšar, "A combinatorial approach to graphlet counting," *Bioinformatics*, vol. 30, no. 4, pp. 559–565, 12 2014.

[34] D. Q. Goldin and P. C. Kanellakis, "On similarity queries for time-series data: constraint specification and implementation," in *International Conference on Principles and Practice of Constraint Programming*. Springer, 1995, pp. 137–153.

[35] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PloS one*, vol. 6, no. 8, p. e23176, 2011.

[36] B. D. Ripley and W. Venables, *Modern applied statistics with S*. springer, 2002.