Time-weighted Social Network: Predict When an Item Will Meet a Collector.

Guillaume Blot and Pierre Saurel
Paris-Sorbonne University
SND FRE 3593
Paris, 28 rue Serpente 75006
guillaume.blot@paris-sorbonne.fr

Francis Rousseaux
Reims Champagne Ardenne University
CReSTIC EA 3804
Moulin de la Housse
51100 Reims

Abstract—"For what else is this collection but a disorder to which habit has accommodated itself to such an extent that it can appear as order?". Unpacking his library, Walter Benjamin describes how a collection is singular [2]. Collections are not unified wholes, but rather chains of undefined objects. Classify, search, arrange or browse collections are personal processes influenced by internal reflexions. Working on figural and non-figural collections, Piaget and Inhelder explain how space and time influence the way a collector looks to his collection [13]. As a result, representing collections is an issue for computer scientists. Here, we propose a time-based method, which considers chronological events and draws a time-weighted graph defining patterns of items. We therefore show how this graph outputs different results depending on when it is requested. This work is based on an architecture, designed by Openrendezvous.com, a collaborative web-based application helping to make appointments. Our goal is to adapt a social graph used to define the perfect moment for two people to meet, to the collection case. We discuss how we can build a structure that helps to compute the ideal moment for an item to meet a collector.

I. BACKGROUND

A. Introduction

Here we present an attempt to implant in the collection case, a time-based analytical structure designed in the context of professional agendas. Collector case is interesting in the sense that it invokes a wide and general field of related data. This approach considers two domains apparently far away one from another, but that we will try to compare using our analytical structure. In its original context, the model is watching for patterns in series of appointments between two entities. If successful, our approach would be in a position to identify some correlated items in a collector experience and to suggest metrics based on time among these correlated items. A comparative study about other temporal networks is given in section II. Section III and IV show how the graph is constructed for both context and how it can be queried. Finally, we will close this paper by a discussion about the relevance of the proposed implantation and we will try to list some of the identified limits and benefits.

But first, we have to introduce the collector case and to say why Social Network Analysis should be used to manage collections at a digital scale. To this end, we essentially borrow models argued by W. Benjamin in the chapter "Unpacking my library" from his book "Illuminations". We also use concepts presented by G. Wajcman, J. Piaget and B. Inhelder. Then, once the general case of collection will be exposed, we will extend the spectrum to the digital context. To this end we will refer to ideas expressed by one of this article author [14]. In "The Emergence of Knowledge in the Secrecy of our Collections", Francis Rousseaux, Alain Bonardi and Benjamin Roadley speak about human collection versus computers formal language. They also work on a modelisation of collectors originary habits, and they have developed a system to carry it out.

<table>
<thead>
<tr>
<th>Collector case</th>
<th>Openrendezvous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collector</td>
<td>User</td>
</tr>
<tr>
<td>Item</td>
<td>Object</td>
</tr>
</tbody>
</table>

Fig. 1. An analogy of meaning between the two experimented cases. Both case has 2 entities, the original model (Openrendezvous) is using the ontology (Object, User). Object is standing for the purpose of the appointment. Here, our particular instance for collection is (Item, Collector)

Introducing collections G. Wajcman gives us a starting statement: "The reason why nobody looks to a collection, is because it is not a whole piece, but a pattern of singular objects, a piece + a piece + a piece " [18]. As computer scientists, we need a clear view of a fact in order to translate it into a class, in an object-oriented manner. Formalism is a difficulty that engineers are facing designing collection. Let us illustrate the point with an example: the book collector. This person, who shows interest in Platos philosophy will probably own an initial set of items, either hardback, pocketbooks, or digital. In order to complete his collection, he will visit a bookstore, where books are classified conformingly to the sellers desires. That can be confusing for the collector, who is wondering where to search for Platos books: in the Philosophical section? Education section? Ancient times? During his search, he shows some interest for other books with nice paperbacks and purchases some, at the same time as a Platos

1http://www.alainbonardi.net/siteRecollection/
book that he actually found. Once he comes back home, he starts to arrange the completed collection in a very personal and satisfying manner, which is pleasing to the eyes and also allows him to retrieve items quickly. If he decided to buy new books on a website, the problem would be rather similar. The difference is that he would be allowed to search for a name, a period, a style or even through the book digitalized content.

Let us extend the outlook and give instances of collection applications. Nowadays we are able to record habits about IT users (either digital and traditional visits of items), we can imagine a system which aggregate heterogeneous information about people habits. Figures 3, 4 and 5 show the minimum information we need to run our model. But, these data can also be found from different sources, such as external API or also Linked-Data ², we can imagine to handle data such as the items that have been visited by a user and the moment when he visited them. For example, an entity such as an event management association or a cultural service of a city hall could be informed on a targeted audience habits and then produce suggestions about what could be their activity timeline, all of this based on people known activities.

B. Collectors and Collections: A complex approach

This subsection aims to give further details on the complex aspects of collections. In current literature, a complex system is a concept in nature or in society, which is clearly hard to emphasise entirely. Complex situations integrate several modular and internal processes which make a whole that can’t be simplified [11]. Typically, the entire system is not equal to the sum of all of its parts. Additional dynamics emerge from the linkage of the finer grain modules. If we consider this definition and the present introduction about collection, we can say that the collector behavior toward his collection is a complex system.

Rendering and exploring complex systems can be done in some ways thanks to large graphs [4] [17] [19]. The question is how to build a Social Graph able to render some part of a collector experience? What can be efficiently measured? Let see now how find a way in this very large entropy.

Walter Benjamin, gives us granular details on collections singularity: "Dates, place names, formats, previous owners, bindings, and the like: all these details must tell him [the collector] something not as dry, isolated facts, but as harmonics whole " [2]. In this first statement, W. Benjamin tells us that collections are personal processes led by internal processes and a various types of reasons. "Books have their own destinies" is a well-known statement, but for W. Benjamin, "also copies of books have their fates". This second statement can be divided in two points. It means that two copies of an item may not be aquired for the same reasons by different collectors. Moreover, it means that a digital and a paperback version of a book would hit the same collector in different ways.

In another environment, Piaget and Inhelder have discussed formal and informal shapes of collections and shows that several steps take place, in someone relationship toward a collection. This relationship changes over time [13].

Our observation here is to say that Time is an important factor influencing the collector behavior.

Fig. 2. Each collector has a personal experience toward his collection. This experience can be recorded by everydaylife tools. This figure could remind timelines used by social networks software like facebook or twitter, which record users activities.

Fig. 3. Here, the (item, collector) relationship is illustrated by a relational database, but it can take many different shapes.

<table>
<thead>
<tr>
<th>id</th>
<th>label</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>American Hustle</td>
<td>movie</td>
</tr>
<tr>
<td>2</td>
<td>SN Analysis(S. Wasserman)</td>
<td>hardback</td>
</tr>
<tr>
<td>3</td>
<td>Blue train (John Coltrane)</td>
<td>CD</td>
</tr>
<tr>
<td>4</td>
<td>Blue train (John Coltrane)</td>
<td>MP3</td>
</tr>
<tr>
<td>5</td>
<td>Georges Braque: A retrospective</td>
<td>exhibition</td>
</tr>
<tr>
<td>6</td>
<td>Conference TXHDC</td>
<td>conference</td>
</tr>
<tr>
<td>7</td>
<td>Magritte: The Mystery of Ordinary</td>
<td>exhibition</td>
</tr>
<tr>
<td>8</td>
<td>Blue in Green (Miles Davis)</td>
<td>CD</td>
</tr>
<tr>
<td>9</td>
<td>Be good (G. Porter)</td>
<td>CD</td>
</tr>
<tr>
<td>10</td>
<td>Museum of fine arts</td>
<td>museum</td>
</tr>
<tr>
<td>11</td>
<td>Imaginated Communities ED 1</td>
<td>hardback</td>
</tr>
<tr>
<td>12</td>
<td>Liquid Spirit (G. Porter)</td>
<td>theatre</td>
</tr>
<tr>
<td>13</td>
<td>LA Woman (The doors)</td>
<td>PDF</td>
</tr>
<tr>
<td>14</td>
<td>Hallucinations (O. Sacks)</td>
<td>pocketbook</td>
</tr>
<tr>
<td>15</td>
<td>Incognito (D. Eaglman)</td>
<td>PDF</td>
</tr>
<tr>
<td>16</td>
<td>Imaginated Communities ED 1</td>
<td>hardback</td>
</tr>
<tr>
<td>17</td>
<td>Illuminations (W.Benjamin)</td>
<td>MP3</td>
</tr>
<tr>
<td>18</td>
<td>Illuminations (W.Benjamin)</td>
<td>movie</td>
</tr>
<tr>
<td>19</td>
<td>Port of Miami (Rick Ross)</td>
<td>movie</td>
</tr>
<tr>
<td>20</td>
<td>Four rooms</td>
<td>pocketbook</td>
</tr>
<tr>
<td>21</td>
<td>Fighters</td>
<td>exhibition</td>
</tr>
<tr>
<td>22</td>
<td>Metamorphosis (F.Kafka)</td>
<td>pocketbook</td>
</tr>
<tr>
<td>23</td>
<td>Vikings: Life and Legend</td>
<td>exhibition</td>
</tr>
</tbody>
</table>

²http://linkeddata.org/
II. STATE OF THE ART

A. Openrendezvous

The model presented here, is based on Openrendezvous.com\(^3\) architecture, a collaborative and social calendar, designed by Guillaume Blot.

Openrendezvous.com use an analytical graph to set appointments. The system is in a position to find patterns of appointments. Our goal is to transpose the appointment social graph into collection case. For example, if a dentist is taking an appointment with a patient, the system helps to determine what is the next sequence of appointments and when the appointment should be set, regarding to the object. All of this is based on historical data recorded among the overall system. The reason why a period is separating two appointments, is because the patient needs a one week antibiotic treatment. The system doesn’t have to know that pulling a teeth involves such a treatment. It just has to check in its historical section, the duration range between two objects. Famous efficient computer-generated recommendation systems have been built using correlations between data, such as the one developed by Greg Linden, followed by thousands of web sites \(^12\).

The system can predict some parts of users future behaviors just by considering historical connections between objects. No matter why it is linked, the fact that it is connected in some way, can be exploited. Predictions models are essantially based on historical data \(^21\) \(^15\).

Based on social network analysis, we propose a solution to place a collector in a singular mechanism, helping him to classify, arrange, and browse his own collection, independently of the chosen classification model. The solution is a social network, in the top-layer of an information system, where edges are valuated with a time-based measurement. This approach use the lifestream metaphor, which aims to schedule people tasks depending on their previous digital life \(^7\).

B. Analytical processes based on time

Some temporal networks already exists. Major recent models have been demonstrated based on an experiment led in different social contexts. The experiment consists in periodically capture a group movement, with proximity sensors installed on persons. The data where collected by SocioPatterns collaboration\(^4\). Based on these data, L. Gauvin et al. used an epidemic process to measure the spreading of information. The methodology consists in giving a piece of information to an entity and measure the moment when the information is potentially possessed by all entities of the system. They compare the spreading dynamics among the recorded human contacts and by creating other random temporal graphs. This comparative study proves the importance of some elements of the topology in the spreading dynamics \(^9\).

Based on the same kind of data, C. Cattuto et al. propose a representation of the system, using NEO4J\(^5\) tool. The result is an animated graph where each pulse corresponds to a frame. This approach is interesting in the sense that it records events and places it on a timeline. Many information can be found by querying the structure. For example, it is possible to ask for the presence of people given several frames or to get the weighted-proximity during a given frame. A recent publication presents their work and gives performance details in terms of access time \(^3\).

Analysing dynamics and network evolution, both initiatives provide major improvement for the field of longitudinal networks. Through the use of longitudinal networks, we plan to render evolution, such as nodes removal or metrics fluctuation. Typical studies related to longitudinal networks, present few isolated representations over a large period and start an analysis \(^6\). Innovation clearly appears in the fact that these two initiatives give tools to arrange and browse a graph over a period, with just one representation.

Besides, the solution that we present in the following sections has different purposes. As a consequence the topology diverges in many ways. We can see in Figure 6, major differences between our model and the two that we just presented. The most significative difference is the number of node classes. Our model is a one-mode network, which is simpler to analyse. Therefore it orients the purpose. Our model purpose is to find sequences and deliver a temporal measure within this sequence, whereas the Cattuto et al. model retrieves information in a more general way.

<table>
<thead>
<tr>
<th>Cattuto et al.</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td>graph amount</td>
<td>n</td>
</tr>
<tr>
<td>nodes</td>
<td>m: collector, frame...</td>
</tr>
<tr>
<td>applications</td>
<td>query</td>
</tr>
</tbody>
</table>

Fig. 6. This table compares some elements of topology between the representation of Cattuto et al. and our representation. Orignial terminology has been replaced in order to adapt the general orientation to the collector case. For example, the Collector entity is originally called Actor.

III. SOCIAL NETWORK ANALYSIS

A. Our Time graph topology

In this section we are giving details about our time-weighted graph topology. The following section (IV) shows some ex-

\(^3\)http://www.openrendezvous.com  
\(^4\)http://www.sociopatterns.org  
\(^5\)http://www.neo4j.org
ploitations that could be done using the presented architecture.

Social network analysis is based on graph theory. In this context, using a graph implies particular considerations. Depending on the number of node classes, we define the social graph mode. If we want to use analysis methods, the social graph should not go further a two-mode topology. In this approach, a node is an item: we have just one class (Figure 4: an item is a movie, a music or a book). Hence, we don’t need to use projection and clustering methods [19] [20] [16].

- We define relationship rules: two nodes are connected, if they have a direct chronological relation. These two nodes are interlocked. Figure 7 shows a chronological chain of items, acquired by a single collector. He was in relationship with item “1” directly before he was in relationship with item “4”. Furthermore, “1” is chronologically visited before “7”, but because item “4” visit happened meanwhile, it is “4” that is connected to “7”.

- We define the graph weight: depending on what is recordable, one can link entities through various types of interdependant criteria including: style, content, shape, media, date, previous owners or location... Here we plan to build a time-based graph.

![Fig. 7. A single-collector chronological chain of the visited items.](image)

In our approach, we want to handle only one graph by giving to vertices a time-based value. The mechanism consists to measure a period between the use of two items. We extract temporal information from the operational database: the timestamp field in Figure 1, and we build a chronological chain of a user collection. This temporal piece of information is used as vertices weight. In Figure 3, “1”, ”4” and ”7” items refers to the item field from the ITEM table. Edges are valued with the time elapsing between the two items visits. Typically: a collector listens to a streaming music at 3:00PM (i), and then buys a movie ticket at 4:00PM (j). The resulting measure will be:

\[ t(i, j) = 60 \text{mn} \]

B. A multi-collector data model

Once we have introduced a single-collector topology, we can start to build an overall social graph, taking into account all users actions. In this overall configuration, two nodes are connected, if at least one collector has been in a direct relationship, with both dedicated items. Now, as we consider all users, we introduce a new feature: a relationship can be set toward the two directions. The result is a bidirectional time-weighted graph. In one sentence: it is possible for two nodes to be connected in both directions, one direction or not connected at all.

![Fig. 8. Multi-collector time-weighted topology.](image)

(a), means that at least one collector has visited both items, visiting item ”1” in the first place, but no user has used both items, visiting item ”2” first.

(b), means that at least one collector has used both items, visiting item ”1” in the first place, and at least one other collector has used both items, visiting item ”2” first.

We define how to calculate a multi-collector vertex value: an edge is weighted by the average time between the visit of ”i” and the visit of ”j”. The following formula computes the weight of an edge. "p" stands for the number of nodes that have directly visited ”i” before ”j”:

\[ t(i, j) = \frac{\sum_p t_p(i, j)}{n_p}, (t(i, j) \neq t(j, i)) \]

In a time-weighted social graph, there is no specific need of the shortest path algorithm. Internal items patterns that we are searching for, are duration dependants. For example, we can search for patterns lasting one hour, in order to know what could be a collector concerns, one hour from now. Calculating the quickest path is not one of our goal. Moreover, such a query will return all one hour paths, even the ones used by only one collector. Every path has the same significance level: the edge weight value does not take into account the number of time collectors have visited items. Therefore, \(t(i,j)=30\text{mn}, \text{with } p=100\) and \(t(i,k)=30\text{mn}, \text{with } p=1\) have the same significance level.

What can we do to handle this issue? If we want to find paths with strong importance, one can use complex weighted principles [1]. In our social graph, there is a way to take into account the importance of nodes, and then to favorise a path over others: nodes degree. The criterion used to calculate the degree is the number of adjacent vertices (more connected nodes are more central). Indeed, we can also consider the number of in or out vertices. Moreover, for reasons explained above, that cannot be completed with the strength, which is the total weight of adjacent vertices. Deeper researches have
to be done here, because significance of the strength value, could fluctuate depending on the context. A single vertex node can have a bigger strength, than a node with ten adjacent vertices. And again, for the same reason as for the shortest path algorithm, here node strength has no such a power on time.

Fig. 9. A time-weighted social network, based on entries from Figure 3, 4 & 5. More the vertex is thick, longer is the interval separating two nodes. We can see that from any starting point, one can chose any existing path and go through the graph.

IV. QUERYING THE GRAPH

Following queries are adapted to the resulting time-weighted social graph topology illustrated by Figure 5. As a matter of fact, queries should output different results in accordance with the current time, and with the considered collector.

Querie 1: What is the first following visited item next The museum of fine art? This querie is simply resolved by returning the smallest *Museum of fine art* output edge.

Querie 2: Build a 20 songs playlist for a specific collector to listen to at 9PM? Suppose the collector last play was Blue train (John Coltrane) at 1PM. Starting at the node “3”, the system inspects the graph for an 8 hours path (9PM-1PM). If the final node is a music item, the first song of the playlist is found. Then, the algorithm will check through the graph, which music must be played next. Typically: find a path that is equal to the duration of the previous track.

Querie 3: Measure the moment when a person who attended to the TXDHC conference has collected the book "Social network analysis". Simply check if there are connections between the two items. Return the weight for both direction.

Querie 4: Return a list of collectors interested in listening to Gregory Porter tomorrow. Browse the graph, searching for a 24hours path, finishing with a Gregory Porter music node. From these resulting paths, make a list with all starting nodes. Any user who is currently visiting a node from the list must be returned.

V. DISCUSSION

Our point is to present an adjustment of an existing model. We are not able to give concrete execution of the engine, but the fact that a concrete model has been done with Openrendezvous.com architecture, helps us to legitimate our researches. Data-model, topology, queries, all of this is based on a concrete architecture. We give this presentation, in order to start a discussion about possibilities offered by the time-based social graph. Firstly, in a general way. And secondly in the particular field of collections, which is much more complex than the simple agenda case. We are aware that the model could have flaws and could be optimised in various points, either in the field of appointments and in the field of collections.

Openrendezvous.com uses appointments as nodes. In an agenda case, an appointment has less material considerations than collectors. Well, indeed, pulling a teeth needs a plier and a teeth. But, the process is always the same: the dentist has a plier, the patient has a teeth, and needs to have a one week prior antibiotic treatment. Here, we can easily draw a model. The object can be reproduced and can take place in different locations. It is not working the same for collections, which can have limited occurrences and not recurrent items. One can notice that these cases have not been addressed in this paper example. A typical instance of a limited occurrence could be an author manuscript, and a none recurrent item could be a concert. However, in a digital system, reproduction concept is redefined.

Another approach could be to draw an affiliation network. Affiliation networks are two modes networks including actors and events. This method is very interesting if we wish to discover memberships among people[20]. Here, relationship mesurement can be anything: frequencies, dates or kind of relation. In our agenda case, the key would be to mesure
edges with time-based value. Instances of affiliation networks are directly or indirectly used in social science [10] [5]. In the context of Openrendezvous, appointments would be events. In the collection context, events would be items and actors would be collectors. This kind of graphs differs from traditional ones. As a consequence it makes the analysis somewhat different than one-mode network. As a result, the solution presented here implements a one-mode architecture.

Besides this issue, the main strength of this work, is to show a mechanism able to classify and browse items, depending on a specific moment. Difficult elements we had to deal with, was singularity and the huge amount of parameters. The "time" parameter is just one aspect, among a whole list of criteria influencing classification. Further work shall consist in designing the graph in a crossing criteria manner. Of course, giving a complete collection model, could not be a one paper work. But, it helped us to emphasis benefits offered by the time-based graph. We welcome feedback, critiques and further suggestions, in order to improve the model.

Future works will be a more in-depth focus on the graph topology and measures. As the graph grows, we will be able to collect new data and to run more specific algorithms. Data are to be collected in the fields of E-learning and Collection.

REFERENCES