



HAL
open science

Delta-Duplication of Time-Varying Graphs

François Queyroi

► **To cite this version:**

François Queyroi. Delta-Duplication of Time-Varying Graphs. Conférence MARAMI 2014 (Modèles et Analyses Réseau: Approches Mathématiques et Informatiques), Oct 2014, Paris, France. hal-00996362v2

HAL Id: hal-00996362

<https://hal.science/hal-00996362v2>

Submitted on 20 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Δ -Duplication of Time-Varying Graphs

François Queyroi

Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris
CNRS, UMR 7606, LIP6, F-75005, Paris, France
francois.queyroi@lip6.fr

ABSTRACT. We introduce a transformation of time-varying graphs, called Δ -duplication, that reduces the temporal heterogeneity in the context of dynamic networks analysis. Instead of building a sequence of snapshots from a global time partitioning, we propose a individual-centred approach: we duplicate a vertex for every time periods where the vertex interacts at least every Δ . This short note describes the general theory of Δ -duplication and provides some directions for applications to dynamic networks analysis. In particular, we introduce a generalization of the concept of k -cores to temporal graphs using this model.

RÉSUMÉ. Nous présentons une transformation de graphes temporels, appelée Δ -duplication, permettant de réduire l'hétérogénéité temporelle dans l'analyse de réseaux dynamiques. Au lieu de construire une séquence d'instantanées à partir d'un découpage global du temps, nous utilisons une approche centrée sur les individus en considérant un sommet sur plusieurs sessions i.e. des périodes durant lesquelles il se connecte au moins tous les Δ . Cette note décrit précisément le concept de Δ -duplication et fournit des pistes quant à son utilisation pour l'analyse de réseaux complexes. En particulier, nous proposons une généralisation du concept de k -cores aux graphes temporels.

KEYWORDS: dynamics networks, graph mining, k -cores

MOTS-CLÉS: réseaux dynamiques, fouilles de graphes, k -cores

DOI:10.3166/DN...1-?? © 2014 Lavoisier

1. Introduction

Complex networks often correspond to dynamic systems such as communications networks whose changes can be tracked in almost real time. In this context, time-varying graphs (Casteigts *et al.*, 2011) (denoted TVG) are used to model this dynamic. The interactions between actors in the network appear and disappear according to an unknown meta-structure (e.g. the existence of *communities*). This process is heterogeneous, in the sense it may evolve over time (the communities change). Dynamic networks analysis aims at the understanding of this evolution in order to make predictions or detect events.

One way to deal to reduce the temporal heterogeneity by splitting the time into different windows of fixed length h . The objective is to discard “short-term” fluctuations and focus on the “long-term” evolutions. This approach transforms a TVG into a series of static graphs called *snapshots* (Hopcroft *et al.*, 2004; Leskovec *et al.*, 2005). An example can be found in Figure 1(a) where a temporal graph with 6 vertices is transformed into two snapshots. Even though this transformation has an important effect on the analysis, it is rarely discussed in the literature.

An implicit hypothesis is that the way interactions are ordered in those snapshots is not relevant for the analysis. The snapshots can then be studied as static graphs (eventually weighted). So far this reduction is done globally (*e.g.* by considering all interactions within an hour, a day, a week) using time windows which seems relevant for the applications’ domains. Such “vertical” time cuts may breaks some relevant patterns. In Figure 1(a), the transformation breaks the triangle formed by vertices (d, e, f) . Notice it is less likely if the cut occurs when the network activity is low (*e.g.* during the night in a human interactions network). The choice of the windows length is therefore not the only issue with the snapshot transformation. There is no guarantee that the time cuts always correspond to inactivity periods regardless of the windows length.

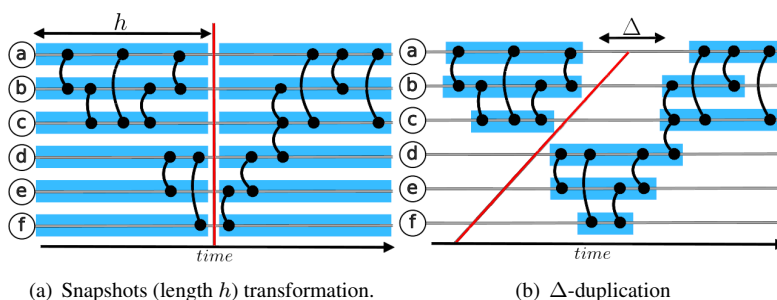


Figure 1. Illustration of the difference between snapshots and Δ -duplication on a time-line representation of a TVG. Bended links are the temporal edges. The blue boxes represent the different instances of a vertex that are created. The red lines represent the induced time cuts.

This preliminary work focuses on the design of an alternative and more general data transformation that does not use an absolute time-line. We propose to partition each vertex time-line into *sessions* according to a parameter Δ : time intervals in which an individual interacts at least every Δ (see Figure 1(b)). This gives birth to a class of TVG we call Δ -successive and we show how a TVG can be turned into a unique Δ -successive TVG (see Section 2). We discuss in what extent the transformed TVG could be studied as a static graph. This transformation allows us to define an important network metric, the k -cores, in the context dynamic networks (see Section 3). The (k, Δ) -cores can be used to detect temporal subgraphs with low-bounded temporal and topological connectivities.

2. Δ -Duplication of Temporal Graphs

We call a TVG a tuple $G = (V, E, T, t)$ where V is the set of vertices, $E = V \times V$ the set of edges and $t : E \rightarrow [0, T]$ is the time at which an edge is observed. We only consider instantaneous interactions. However, most of the concepts developed later can be defined for more comprehensive models (see Section 4). We call $t(v)$ the set of different timestamps for which it exists an edge incident to $v \in V$. Furthermore, the interval $[\max t(v), \min t(v)]$ is called the *session* of v (activity period).

DEFINITION 1. — *Δ -successive Graphs.* A temporal graph $G = (V, E, t)$ is said to be Δ -successive for $\Delta \in [0, T]$ iff for all $v \in V$, $t(v)$ is a Δ -session i.e. the biggest interval between two consecutive interactions in $t(v)$ is lower or equal to Δ .

Testing whether a temporal graph is Δ -successive is straightforward. We can derive from this definition some useful properties.

PROPERTY 2. — *Minimum frequency.* If G is Δ -successive, the frequency of interactions of $v \in V$ when active is $d_G(v) / (\max t(v) - \min t(v)) \geq \Delta^{-1}$

PROPERTY 3. — *Inclusion.* For $\Delta_1 \leq \Delta_2$, every Δ_1 -successive graph is also Δ_2 -successive.

For every TVG it exists a Δ such that the graph is Δ -successive. We show now that any TVG can be turned into a Δ -successive with lower Δ using vertex duplication.

DEFINITION 4. — *Vertex-Duplication.* The graph $S = (V', E, \theta)$ is a Vertex-duplication of $G = (V, E)$ iff $\theta : V' \rightarrow V$ is a surjective function such that the graph obtained after the contraction of the vertices $\{v \in V', \theta(v) = u\}_{u \in V}$ is G .

A vertex-duplication is a transformation of G without information loss. In the Δ -duplication, a vertex is seen as separate entities if the time spent between two consecutive interactions is higher than Δ . The function θ can be viewed as a label keeping track of the identify of the original vertex. In Figure 1(b), each blue box corresponds to a Δ -session, the vertices (a, b, c) are duplicated.

DEFINITION 5. — *Δ -duplication.* The temporal graph $S = (V', E, T, t, \theta)$ is a Δ -duplication of $G = (V, E, T, t)$ iff S is a vertex-duplication of G and is Δ -successive.

We call minimum Δ -duplication of G the Δ -duplication where $|V'|$ is minimum. Notice that the minimum T -duplication of G is itself. Equivalently, the minimum 0-duplication of G contains $\sum_{u \in V} |t(u)|$ vertices. It is easy to show that this minimum is unique. Indeed, it corresponds to the partition of each set $t(v)$ into the largest Δ -sessions. This is achieved by splitting only the intervals of length greater than Δ between consecutive interactions.

Computing the Δ -duplication therefore requires the sorting of the edges of G which can be done in $\mathcal{O}(|E| \log |E|)$ (reduced to $\mathcal{O}(|E|)$ if we assume the edges are already sorted). Moreover, the Δ -duplication can be stored in $\Theta(|E|)$ since the edge set is the same as G and $|V'| \leq 2|E|$ assuming $|V| \leq 2|E|$.

We discuss now the usage of a Δ -duplicated TVG as a static graph. The Δ -duplication transforms a TVG into another TVG. As for the snapshots transformation, the assumption can be made that the time ordering of the links is not relevant after this transformation.

One interesting aspect of Δ -duplication is that temporal properties can be turned into topological properties. For example, if it exists regular time periods where the network activity is null (during the night for example). Then for some value of Δ lower than this duration, the activity periods will correspond to disconnected subgraphs (in the static sense) in the Δ -duplication. In this context, the connected components of the transformed TVG can be viewed as a snapshot sequence. However, since Δ -duplication is vertex-oriented, the time cuts may correspond to non-vertical time periods (as in Figure 1(b)). This is useful when the actors are not active during the same periods. This can be the case for worldwide networks where actors belong to different time zones.

3. Generalization of k -cores to temporal graphs

The k -core decomposition is a powerful tool for network analysis (Seidman, 1983). It assigns to each vertex the largest k for which the vertex belongs to a k -core (see Def. 6). Previous studies (Kitsak *et al.*, 2010) suggested that the latter statistics is positively correlated to the ability of the vertices to spread information or disease effectively. It is therefore relevant to generalize this metric to dynamic networks.

DEFINITION 6. — *k -core.* The k -core of a graph $G = (V, E)$ is the maximal subgraph of G with minimum degree at least k .

The idea of the k -core for TVG, called (k, Δ) -cores, is to extract subsets of interactions that form Δ -sessions containing a given minimum number of connections. Our generalisation preserves the properties of uniqueness and inclusion (see Theorems 8 and 9).

The parameter Δ provides a bound to the temporal connectivity of a TVG (see Property 2). The parameter k in the k -core decomposition provides a bound to the topological connectivity (the degree). We now bring those two constraints together. To do that, observe the definition 6 can be formulated in term of edge subsets, which is useful here since the vertices set changes with Δ while the edge set of a TVG stays the same. The concept of k -core can also be defined for various vertex statistic like the in or out-degree or the number of triangles. Since temporal graphs are generally multiple, a relevant statistic could here be the number of distinct neighbours.

DEFINITION 7. — *(k, Δ) -core.* Let G be a TVG, the (k, Δ) -core of G denoted $C_{k, \Delta}(G)$ is the maximal subset of edges such that the subgraph formed by $C_{k, \Delta}(G)$ has a Δ -duplication where the minimum degree is at least k .

THEOREM 8. — *Uniqueness.* For every temporal graph $G = (V, E, T, t)$, it exists a unique (k, Δ) -core.

PROOF. — Suppose it exists two different (k, Δ) -cores of G denoted $C_1 \subset E$ and $C_2 \subset E$ with $C_1 \neq C_2$. Since both C_1 and C_2 are maximal, the graph formed by the union $C_1 \cup C_2$ should not have a Δ -duplication with a minimum degree of k . For a vertex $v \in V$, its incident edges are therefore split into two groups : $C_1(v)$ and $C_2(v)$. Both of these sets can be partitioned into Δ -sessions having at least k elements. Observe that merging a couple of Δ -sessions in $C_1(v)$ and $C_2(v)$ that overlap over their time periods produces a larger Δ -session. Therefore, doing the union of the two sets $C_1(v)$ and $C_2(v)$ and merging the pairs that overlap produce a set of Δ -sessions each having at least k elements. Since it is true for every vertices, it invalidates our hypothesis. We conclude it exists no (k, Δ) -cores C_1 and C_2 such that $C_1 \neq C_2$. ■

THEOREM 9. — *Inclusion. For $\Delta_1 \leq \Delta_2$ and $k_1 \leq k_2$, we have*

$$C_{k_2, \Delta_1}(G) \subseteq C_{k_1, \Delta_2}(G) \quad (1)$$

PROOF. — According to Property 3, a Δ_1 -successive graph is also Δ_2 -successive. For all $\Delta \in [\Delta_1, \Delta_2]$, the edge set $C_{k_2, \Delta_1}(G)$ has a Δ -duplication with minimum degree at least k_2 , therefore it has one with minimum degree at least k_1 . ■

Theorem 9 indicates that a partial order exists between the different cores. It means that one can find the largest k for a non-empty (k, Δ) -core when Δ is fixed. Equivalently, one can find the smallest Δ such that a (k, Δ) -core is non-empty for a given k .

Algorithm 1: Computation of (k, Δ) -cores

Input: $G = (V, E, T, t)$, k , Δ
Output: $C_{k, \Delta}(G)$

- 1 $G' \leftarrow G$;
- 2 **while** $\min_{u \in V(G')} d_{G'}(u) < k$ and G' not Δ -successive **do**
- 3 $G' \leftarrow \text{min-}\Delta\text{-duplication}(G')$;
- 4 $G' \leftarrow k\text{-core}(G')$;
- 5 **end**
- 6 **return** $E(G')$;

Algorithm 1 can be used to compute the (k, Δ) -core of a temporal graph. One difficulty is that a subgraph of a Δ -successive graph may not be Δ -successive. Therefore the peeling of vertices with low degree (line 4) must be used in conjunction with the computation of minimum Δ -duplications (line 3) until the conditions given in Definition 7 are met. Finding the k -core of G' (considered here as a static graph) can be done in $\mathcal{O}(|E|)$.

THEOREM 10. — *Correctness of Algorithm 1. For a TVG $G = (V, E, T, t)$, the Algorithm 1 returns $C_{k, \Delta}(G)$.*

PROOF. — We call degree (resp. core value) of an edge in G the minimum degree (resp. core value) of its endpoints. First, at each iteration of the while loop, we

have $C_{k,\Delta}(G) \subseteq E(G')$. The transformation into minimum Δ -duplication can only decrease the core values. Those values are not modified for the edges that remain after the extraction of the k -core of G' . Moreover, the Δ -duplication for which the core of values of edges is maximum is the minimum Δ -duplication of G' . For every edge $e \in C_{k,\Delta}(G)$, the core value of e is therefore at least k at each iteration and $C_{k,\Delta}(G) \subseteq E(G')$ at the end of the algorithm. Next, $E(G') \subseteq C_{k,\Delta}(G)$ since G' is a Δ -successive subgraph of G with a minimum degree of k . By definition, the edges of G' belong to $C_{k,\Delta}(G)$. Therefore, we have $E(G') = C_{k,\Delta}(G)$. ■

4. Discussion and Future Directions

Our formalism can be extended to various definitions of time-varying graphs. For example, non-instantaneous interactions can be used *i.e.* when edges correspond to quadruple $(u, t_1, v, t_2) \in V \times E \times V \times E$ or continuous time intervals $(u, v, [t_1, t_2])$. In the later, the edge (u, v) is active during the period $[t_1, t_2] \subseteq T$. The Δ -duplication depends on the notion of vertex activity and we consider a vertex inactive if it is not the extremity of an edge for a duration of at least Δ . This concept is still valid for the TVG definitions given here although different types of computation may be needed.

Our objective is to use Δ -duplication and (k, Δ) -cores in the practical context of dynamic networks analysis. Since our transformation relies on a parameter Δ , we want to design methods to effectively explore the space of Δ -duplication of real-world networks. In particular, computing the (k, Δ) -cores for different parameters values may provide a good fingerprint of a given network. We want to compare the information obtained and the pattern detected by comparing them to the conclusion of studies based on a transformation into snapshots of fixed or varying length.

References

- Casteigts A., Flocchini P., Quattrociocchi W., Santoro N. (2011). Time-varying graphs and dynamic networks. In *Ad-hoc, mobile, and wireless networks*, pp. 346–359. Springer.
- Hopcroft J., Khan O., Kulis B., Selman B. (2004). Tracking evolving communities in large linked networks. *Proceedings of the national academy of sciences of the United States of America*, Vol. 101, No. Suppl 1, pp. 5249–5253.
- Kitsak M., Gallos L. K., Havlin S., Liljeros F., Muchnik L., Stanley H. E. *et al.* (2010). *Identification of influential spreaders in complex networks*. *Nature Physics*, Vol. 6, No. 11, pp. 888–893.
- Leskovec J., Kleinberg J., Faloutsos C. (2005). *Graphs over time: densification laws, shrinking diameters and possible explanations*. In *Proceedings of the eleventh acm sigkdd international conference on knowledge discovery in data mining*, pp. 177–187.
- Seidman S. (1983). *Network structure and minimum degree*. *Social networks*, Vol. 5, No. 3, pp. 269–287.