Finding cohesive communities with C^3

Adrien Friggeri, Eric Fleury

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Finding cohesive communities with C^3

Adrien Friggeri — Eric Fleury

N° 7947

30 April 2012

ISSN 0249-6399 ISRN INRIA/RR--7947--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche GRENOBLE - RHÔNE-ALPES

Finding cohesive communities with C^3

Adrien Friggeri , Eric Fleury

Thème : Réseaux et télécommunications
Équipe-Projet DNET

Rapport de recherche n° 7947 — 30 April 2012 — 27 pages

Abstract: Social communities have drawn a lot of attention in the past decades. We have previously introduced and validated the use of the cohesion, a graph metric which quantitatively captures the community-ness in a social sense of a set of nodes in a graph. Here we show that the problem of maximizing this quantity is NP-Hard. Furthermore, we show that the dual problem of minimizing this quantity, for a fixed set size is also NP-Hard. We then propose a heuristic to optimize the cohesion which we apply to the graph of voting agreement between U.S Senators. Finally we conclude on the validity of the approach by analyzing the resulting agreement communities.

Key-words: graph theory, community detection, np-completeness, cohesion, complexity, social network analysis

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Trouver des communautés cohésives avec C^3

Résumé : Les communautés sociales ont attiré beaucoup d'attention ces dernières années. Nous avions précédemment proposé et validé l'utilisation de la cohésion, une métrique de graphe qui capture quantitativement la qualité communautaire, au sens social, d'un ensemble de sommets d'un graphe. Nous montrons que le problème de trouver un ensemble de cohésion maximum dans un graphe non orienté est NP-dur. Par ailleurs, nous montrons que le problème dual de minimiser cette quantité, pour une taille donnée, est aussi NP-dur. Nous proposons ensuite une heuristique pour optimiser la cohésion que nous appliquons au graph d'agrément de vote entre Sénateurs des États-Unis. Finalement nous concluons sur la validité de l'approche en analysant les communautés résultantes.

Mots-clés : théory des graphes, détection de communautés, np-completude, cohésion, compléxité, analyse de réseaux sociaux

In [1], we have introduced a new metric called the *cohesion* which rates the community*ness* of a group of people in a social network from a sociological point of view. The idea behind the cohesion is, rather than looking at the proportion of edges falling inside and in between communities, to take into account the triads in the network and define a community as a subgraph having a high transitivity and featuring a low number of triangles going outwards to the rest of the network. Through a large scale experiment on Facebook, we have established that the cohesion is highly correlated to the subjective user perception of the communities.

In this article, we show that finding a set of vertices with maximum cohesion is $\mathcal{NP}$-hard. We will then also establish that the dual problem of finding the less cohesive groups of a graph is $\mathcal{NP}$-hard. Then we shall introduce C^3, a heuristic which covers a given graph with cohesive communities by pseudo-greedily expanding around selected edges.

Finally we shall validate this heuristic by studying the communities it yields on the agreement graph of U.S. Senators on which we shall be able to demonstrate that the communities which are obtained independently for each Congress Session are stable through time and can be identified with political parties.

Notations Let $G = (V, E)$ be a graph with vertex set V and edge set E of size $n = |V| \geq 4$. For all vertices $u \in V$, we write $d_G(u)$ the degree of u, or more simply $d(u)$[1] and $\mathcal{N}(S)$ the set of neighbors of S. A *triangle* in G is a triplet of pairwise connected vertices. For all sets of vertices $S \subseteq V$, let $G[S] = (S, E_S)$ be the subgraph induced by S on G. We write $m(S) = |E_S|$ the number of edges in $G[S]$, and $\text{◬}(S) = |\{(u, v, w) \in S^3 : (uv, vw, uw) \in E_S^3\}|$ the number of triangles in $G[S]$. We define $\text{⏏}(S) = |\{(u, v, w), (u, v) \in S^2, w \in V \setminus S : (uv, vw, uw) \in E^3\}|$, the number of *outbound* triangles of S, that is: triangles in G which have exactly two vertices in S. Finally, we recall the definition of the cohesion.

$$\mathcal{C}(S) = \frac{\text{◬}(S)^2}{\binom{|S|}{3}(\text{◬}(S) + \text{⏏}(S))}$$

The cohesion is a measure of the community-ness of a set of nodes and is a compromise between a large density of triangles inside the community and the amount of triangles pointing outwards from the community.

1 Complexity

1.1 Max-Cohesion is $\mathcal{NP}$-hard

In this section we examine the problem of finding a set of vertices $S \subseteq V(G)$ of maximum cohesion in a graph G, i.e. for all subset $S' \subseteq V$, $\mathcal{C}(S') \leq \mathcal{C}(S)$. We shall first show that the set of vertices with maximum cohesion in a given network is connected, per Theorem 1.2.

[1] Here, as elsewhere, we drop the index referring to the underlying graph if the reference is clear.

Lemma 1.1 *Let $S_1, S_2 \subseteq V(G)$ be two disconnected sets of vertices ($(S_1 \times S_2) \cap E(G) = \emptyset$). Then $\mathcal{C}(S_1) \leq \mathcal{C}(S_1 \cup S_2) \Rightarrow \mathcal{C}(S_2) > \mathcal{C}(S_1 \cup S_2)$. The proof is given in A.1.*

Theorem 1.2 *Let S be a non-connected set of vertices of G. Then there exists a connected set $S' \subseteq S$ having a higher cohesion $\mathcal{C}(S') > \mathcal{C}(S)$.*

Proof *Let $S_1, S_2 \subseteq V(G)$ such that $S_1 \cup S_2 = V(G)$ and S_1, S_2 disconnected, then at least one of S_1 or S_2 has a higher cohesion than S per Lemma 1.1. If the set with higher cohesion is connected, the result is immediate. If not, the same reasoning applies to that set, which leads to the conclusion.* □

Therefore the problem at hand is equivalent to that of finding a connected set of vertices with maximum cohesion in G. The decision problem associated to the latter is CONNECTED-COHESIVE.

Input	A graph $G = (V, E)$, $\lambda \in \mathbb{Q}$, $\lambda \in [0, 1]$
Question	Is there a subset connected S of V such that $\mathcal{C}(S) \geq \lambda$?

We shall now proceed to show that CONNECTED-COHESIVE is $\mathcal{NP}$-complete. First note that given a set S of vertices of G, it is possible to verify that S is a solution of CONNECTED-COHESIVE by computing its cohesion, its size, its connectivity and the minimum degree of its vertices, all in polynomial time. Therefore CONNECTED-COHESIVE is in $\mathcal{NP}$. We shall now reduce CLIQUE to CONNECTED-COHESIVE. We recall that CLIQUE is:

Input	A graph $G = (V, E)$, $k \in \mathbb{N}, k \leq \lvert V\rvert$
Question	Is there a subset S of V such that $\lvert S\rvert = k$ and the subgraph induced by S is a clique?

Let $(G = (V, E), k \in \mathbb{N})$ be an instance of CLIQUE[2]. We can assume that G is connected (if not, we use the following reasoning separately on each connected component of G). We construct an instance $(G' = (V', E'), \lambda)$ of CONNECTED-COHESIVE by adding an edge between all non connected vertices u and v in G and then linking those two vertices to all vertices in a clique of size $2\binom{n}{3}^4$ which we add to the network, as described in Algorithm B.1 and illustrated by Figure 1.

Theorem 1.3 *There exists a clique of size k in G if and only if there exists a connected group of vertices of G' with cohesion $\lambda \geq \frac{\binom{k}{3}}{\binom{k}{3}+\binom{k}{2}(n-k)}$. The proof is given in A.2.*

Theorem 1.4 CONNECTED-COHESIVE *is $\mathcal{NP}$-complete.*

Proof. Per Theorem 1.3, there exists a clique of size k in G if and only if there exists a connected subset of vertices of G' of cohesion $\lambda \geq \frac{\binom{k}{3}}{\binom{k}{3}+\binom{k}{2}(n-k)}$ and

[2] We consider here that $|G| > 2$ and $k > 2$, although this is not exactly CLIQUE, this problem is clearly $\mathcal{NP}$-complete.

the transformation from G, k to G', λ runs in polynomial time. Thus CLIQUE is reducible to CONNECTED-COHESIVE and the problem is $\mathcal{NP}$-hard. Given that it is in $\mathcal{NP}$, the problem is thus $\mathcal{NP}$-complete.

The associated decision problem being $\mathcal{NP}$-complete, the problem of finding a set of vertices with maximum cohesion is $\mathcal{NP}$-hard. Note that the problem of finding a set of vertices of maximum cohesion containing a set of predefined vertices is also $\mathcal{NP}$-hard, by an immediate reduction.

1.2 k-Min-Cohesion is $\mathcal{NP}$-hard

Dual to the problem of finding a subset with maximum cohesion is that of minimizing the cohesion, which can be useful when trying to identify socially weak subgraphs in highly cohesive networks. We formulate the k-MIN-COHESION problem in the following way:

Input A graph $G = (V, E)$
Output A subset $S \subseteq V$ such that $|S| = k$ and $\forall S' \subseteq V$, $\mathcal{C}(S) \leq \mathcal{C}(S')$

In order to prove that k-MIN-COHESION is $\mathcal{NP}$-hard, we will show that the problem of finding a set of nodes of size k with cohesion 0 is $\mathcal{NP}$-complete. First note that one can check in polynomial time that a set of nodes has cohesion 0, thus the problem is in $\mathcal{NP}$. Now notice that if a set of nodes has cohesion 0, then in particular $\triangle(S) = 0$, which means that S does not contain any triangles. Conversely, if $\triangle(S) = 0$ then $\mathcal{C}(S) = 0$, therefore finding a set S of size k such that $\mathcal{C}(S) = 0$ is equivalent to the problem of finding a triangle free induced subgraph of size k. The property of being triangle free is hereditary: all subgraphs of a triangle-free subgraph is itself triangle free, and is non trivial: there are infinitely many triangle free subgraphs, therefore the problem of finding a triangle free induced subgraph is $\mathcal{NP}$-complete [2].

2 The C^3 heuristic

In this section we introduce a heuristic algorithm, C^3 which operates in three step: COVER, COMMUNITIZE and COMBINE. C^3 covers a network (COVER) with overlapping communities in a network by greedily maximizing their cohesion (COMMUNITIZE), and merges communities which overlap two much (COMBINE).

2.1 Communitize around a node

We present an algorithm specially tailored for the problem of maximizing the cohesion. Suppose that we have a set of nodes S and we wish to add nodes to S in order to potentially find a new set $S' \supseteq S$ such that $\mathcal{C}(S') \geq \mathcal{C}(S)$. The greedy approach is to start by adding to S a node which increases it cohesion. Another solution is to also explore the possibility that adding a node which increases the transitivity of the subgraph might lead to a higher cohesion in the end, if this turns out not to be as successful as hoped, then we can always revert back to S and try with another unvisited node.

We shall say that u is a **candidate** for S if it follows the constraints which are detailed in Algorithm B.2. First the node should be a neighbor of our seed group because of the connectivity of the community with maximal cohesion and we then verify that the node has not already been added to the community. Next, a node is a candidate if we increase the number of outbound triangles of a group without triangles (adding this type of nodes allows us to go deeper into the "heart" of the communities). In all other cases we forbid nodes who do not create any triangles inside the community. The main reason the two previous constraints are added is to be able to deal with the cases where $|S| \leq 3$, in which case S does not contain any triangle and we need to bootstrap a beginning of community. Finally, we mark as candidates the nodes which increase the cohesion and the transitivity.

The next task is to discriminate which of those nodes is the best possible candidate. Given a set of nodes S and its candidates K, we select the node u with the highest $\mathcal{C}\left(S \cup \{u\}\right)$. If all have identical value, we will then select the one which has the highest $\frac{\textcircled{\triangle}(S \cup \{u\})}{|S|+1}$. In case all nodes have still the same values, we shall discriminate using the highest $\overset{\triangle}{\circ}\left(S \cup \{u\}\right) - \overset{\triangle}{\circ}\left(S\right)$. Finally, in last resort, we will pick the node with the highest degree. These two last criteria are here to help bootsrap the community in the first rounds of the algorithm.

We define the Communitize algorithm (Alg. B.3) recursively as follows, given a graph G and a set of nodes S we first establish the list K of nodes of G which are candidates in respect to S. Then, for each node u of K, chosen as previously mentioned, if we have not already visited that node, we compute the best community containing $S \cup \{u\}$. Finally, we return the community with the highest cohesion in the set $\{S\} \cup \bigcup_{u \in K}$ communitize$(S \cup \{u\})$.

Computing the cohesion of a set of nodes has a non negligeable cost, as it is done in $\mathcal{O}(|S \cup \mathcal{N}(S)|^3)$. It is however not mandatory to recompute at each step the cohesion, as we can just track the variations induced by the addition or deletion of a node by keeping count of the number of inbound and outbound triangles and for each node u, the number of inbound (resp. outbound) triangles $\textcircled{\triangle}_u$ (resp. $\overset{\triangle}{\circ}_u$) which would be added if u was added to S. We can then write an updateSet function which adds or delete a node u to S and maintains the correct values of $\textcircled{\triangle}$ and $\overset{\triangle}{\circ}$. Let $\delta = 1$ when u is added and $\delta = -1$ when u is deleted. If the list of neighbors $\mathcal{N}(u)$ are sorted it is possible to bring the complexity of the update down to $\mathcal{O}(\sum_{v \in \mathcal{N}(u)} d(u) + d(v))$.

By modifying the Communitize algorithm in order to create a community which is then modified in place using setUpdate, we obtain an algorithm which optimizes the cohesion by recursively adding sound candidates to an initial seed and always returns the most cohesive group it encounters. The complexity of adding or deleting a node stems from the update step which cost is $\mathcal{O}(\sum_{v \in \mathcal{N}(u)} d(u) + d(v))$. Each node is added and deleted from the community at most once, which leads to an overall worst case complexity complexity of $\mathcal{O}\left(|V|\,|E|\right)$.

2.2 Cover a set of nodes

Although Communitize allows us to expand around a set of nodes in order to obtain a more highly cohesive set of nodes, it is not always desirable to obtain one and only one community. We introduce a second algorithm (Alg. B.5), which uses Communitize to expand around carefully selected nodes of the network. The basic idea is to choose one node u in S, find the best community containing u and one of its neighbors, mark all the nodes in that community as covered and repeat as long as there are nodes which are not covered. We choose to iterate over S by increasing degree as placing low degree nodes first into their communities allows us to more precisely capture the community structure around nodes with a higher degree. The reason why we choose to expand around u and one of its neighbors is that if we only expanded around u some communities would not be detected. Consider for example the four triangles depicted in Figure 2, the algorithm ignore the presence of the middle clique which nodes would have already been covered.

Finally, we have presented in this section an algorithm which places each node of a given set in at least one cohesive community. Alghough the resulting algorithm has a worst case complexity bounded $\mathcal{O}\left(|V|\,|E|^2\right)$, this would be the case when each edge lead to a community containing all the graph without covering any other edge, which is impossible.

2.3 Combine similar groups

We now present the last pillar of C^3, which allows to parametrically control the amount of overlap which is authorized. Let us consider a graph G. After running Cover we obtain a collection of communities $(S_i)_{0\leq i\leq k}$. Suppose that we are provided with a function Ov which rate at which extent two communities S_i and S_j overlap. Furthermore, suppose that we dispose of a maximum authorized overlap $o_{\max}$. We construct a weighted graph Γ where each node u_i corresponds to a community S_i, and where there are edges between two nodes u_i and u_j if and only if $Ov(S_i, S_j) \geq o_{\max}$, the weight of the edge being $Ov(S_i, S_j)$. Once the communities are laid out this way, the problem of finding sets of communities which overlap sufficiently reduces to a problem of “community” detection in the meta-graph.

However, contrary to the graphs we have encountered until now, Γ is a weighted graph, therefore we have to adapt the definition of the cohesion in order to be able to recursively use C^3 to find the meta-communities. There are several ways the definition of the cohesion can be extended to take into account graphs where edges have weights in $[0, 1]$. Basically, it suffices to produce a function which allows to transfer the notion of weights from the edges uv, uw, vw to the triangle uvw. On can use the product of the edges which is intuitive, however when judging the overlap contained inside a triangle, it might be useful to use the maximum of the product of two of the three edges, which adds transitivity to the overlap function. Using such weights on triangles, the definition of the weighted cohesion comes immediately, in the weighted version, ◬ is the sum of the weights of inbound triangles and $\triangle$ becomes the sum of the weights of outbound triangles.

We can then compute the communities of Γ by using Cover and recursively calling Combine on the result. Then, for each meta-community Σ, three cases are possible, either $|\Sigma| = 1$ and the community in Σ is not affected, either $|\Sigma| = 2$ and the two communities S_1, S_2 in Σ are merged, or $|\Sigma| \geq 3$, the communities in Σ are merged if the cohesion of Σ is higher than a certain treshold $\mathcal{C}_0$. This algorithm always finishes, given that for a graph of size n Cover gives at most $n - 3$ communities.

3 Application of C^3 to the U.S. Senate

The United States Senate is the upper house of the United States legislature. Originally, Senators were elected by the individual state legislatures, but have been elected by the people since the passage of the Seventeenth Amendment in 1913. Contrary to the House of Representative which seats are up for election every two years, Senators serve terms of six years each. Those terms are however staggered so that approximately one-third of the Senate is renewed every two years. This period of two years is called a *United State Congress*.

Contrary to other countries, data concerning elected officials and the activity of the houses is openly available in the United States. We have used the GovTrack website which provides both a list of all elected officials and votes both at the Senate and the Congress to construct graphs of agreement. Each Senator usually serves, except for unfortunate events, in at least three consecutive Congresses. Therefore we have decided to focus on the Senate as there is a continuity in those serving, which allows to observe more precisely the evolution of political groups.

For each of the 112 Congresses we shall construct an agreement graph $G_i = (V_i, E_i)$ where V_i is the set of the Senators active during the i^{th} Congress. Due to some ambiguity in the data, we could not restrict ourselve to the Senators and actually construct the graph of those who have casted at least one vote in Senate, nevertheless, we shall qualify our actors of Senators, for clarity's sake. All votes we have encountered were choices between two options which we shall arbitrarily denote A and B, therefore each Senator had either voted A, or B or did not vote.

For each Congress i we associate to the Senator s a vote vector $V^{i,s}$ of dimension the number of votes which have taken place during that Congress, such that $V_k^{i,s} = 1$ if s voted A for the k^{th} vote, $V_k^{i,s} = 0$ if s did not vote for the k^{th} vote and $V_k^{i,s} = -1$ if s voted B for the k^{th} vote. We can then compute the agreement (or weight) between two Senators as the cosine similarity between their votes $W_i(s_1, s_2) = \frac{V^{i,s_1}.V^{i,s_2}}{\|V^{i,s_1}\|\|V^{i,s_2}\|}$ Given those agreement weights, we can now construct the edges of the agreement graph G_i, we add an edge of weight $W_i(s_1, s_2)$ between s_1 and s_2 if $W_i(s_1, s_2) \neq 0$.

The cumulative distribution of those weights are given, as an example, on Figure 3. It is notable that in all cases more than 50% of the edges have a positive value. We can observe that the earlier Senates presented a certain balance in the distribution of the edges: there was a similar number of edges of positive and negative weights. More recent Senates have a bias towards agreement, as exemplified by the latest Senate (112th Congress (2011–2013)) where 75% of edges have a positive weight and 45% of edges have a value greater than 0.5. This trend is explicit when we observe the evolution of the average value of agreement (Fig. 4). The first thing to notice is that the average value is always greater than 0, which indicates that although being from different political horizons Senators tend to agree more with each other than to disagree. There are however variations in the evolution of the average agreement.

During the first few Congresses, the average agreement increases in a context where the United States are a young nation. There is however a sudden drop in agreement during the Eleventh and Twelfth Congresses, which took place just before and during the war of 1812, the first major conflict between the United States and the British Empire since the end of the American Revolutionary War in 1783. During the Fifteenth Congress (1817–1819), the average agreement rises to more than 0.2. Coincidentally, in 1819, the United States faced the so-called "Panic of 1819", its first major financial crisis. It then steadily decreases, attaigning its minimum during the 27th Congress (1841–1843), in the years of instability leading of the Civil War and then decreasing in average throught the Reconstruction era. It is only with what Mark Twain dubbed the "Gilded Age" that the average agreement increases again, around the time of the 51st Congress (1889–1891). The next major increase occurs around the 83rd Congress (1953–1955). In 1953, major political changes occur both in the United States and the USSR which shifted the dynamic of the cold war. At the same time, Joseph McCarthy started its communist witch hunt while heading the Senate Permanent Subcommittee on Investigations. The Cuban Missile Crisis occured during the 87th Congress (1961–1963) which marked a temporary decrease in agreement, although the average increase would then continue until the 102nd Congress (1991–1993). In the past two decades, the agreement has swinged up and down, staying on average higher than 0.18. Notice how it has attained it has peaked at its maximum during the 107th Congress (2001–2003), which was marked by the 9/11 attacks.

4 Signed & Weighted Cohesion

The graphs G_i that we have obtained in the previous section have weights which vary between -1 and 1, therefore we need to extend the definition of the cohesion in order to take those negative edges into account. If an edge uv has a negative weight, it means that u and v are in disagreement and should not be added to a same community.

In terms of triangle, the consequence is that if a triangle contains at least a negative edge, then it should contribute negatively to the cohesion. We therefore introduce a the $\operatorname{sgn}(uvw)$ function which gives us the sign of the contribution of a triangle, that is $\operatorname{sgn}(uvw) = -1$ if $W(uv) < 0$ or $W(uw) < 0$ of $W(vw) < 0$, and 0 in all other cases. From there we can define the signed weight of a triangle $W_s(uvw) = \text{sgn}(uvw)W(uvw)$, where $W(uvw)$ can be any unsigned triangle weighing function. Here we shall choose to use the product of the edges weights, $W(uvw) = W(uv)W(uw)W(vw)$.

Let us now extend the cohesion in order to take into account the signed weights of triangles and at the same time remain compatible with its unsigned version. If a group has negative △⃝ and positive ⏏, it means that there is more disagreement inside the group than towards the rest of the network, and therefore the cohesion should be low. For similar reasons, if △⃝ if positive and ⏏ is negative, the group has a high agreement with itself and is opposed to the rest of the network, which should result into a high cohesion. Intuitively, if △⃝ and ⏏ are of opposite signs, the group is isolated from the rest of the network and the cohesion is reduced to the transitivity. Finally, there is the case when both △⃝ and ⏏ are negative. In that case the expression of the isolation factor and thus that of the cohesion remain the same. The formulas for the signed and weighted cohesion are given in Table.1. This new definition of the cohesion can be used directly in C^3 without adapting the algorithm.

5 Of History, Dynamics and Stability

For each Congress we have computed using C^3 the communities of its agreement graph using the extended cohesion. On Figure 5 we present the evolution of the number of communities of agreement through time.

The first thing to notice is that, except in three cases, there are between one and three communities. The First Congress (1789–1791) has 10 different communities and the Second Congress (1791–1793) as well as the 37th Congress (1861–1863) has 5 communities. Notably, the latter coincides with the beginning of the American Civil war in 1861 and the larger number of communities reflects the political turmoil at the time. About the two first Congresses, one has to bear in mind that there were no national political parties prior to the Presidential Election of 1796. The United States were a young nation and did not have a two-party system. It is important to note that that era is more an era of faction rather than parties, and thus alliances would shift at a fast pace in this early era of U.S. political history, which is visible when looking at the number of communities during the four first Congresses. By the start of the Fifth Congress, two national political parties had emerged from the two aforementioned factions. From there on, the United States have had a two-party system, which is visible in the number of communities in the agreement graph which except the three previously mentioned exceptions vary between 2 and 3. Furthermore, since the 84th Congress (1955–1957), there was no more than two communities – and there were five occurences where there was only one community. This diminution in the number of communities is a direct consequence of the previously mentioned increase in agreement among Senators.

Concerning the dynamics of those graph, as we have said earlier, at each Congress, only a third of the Senate seats are up for election. On Figure 6, we have plotted the proportion of the members of the Senate during a given Congress which remained in position during the following Congress. This continuity C_i is expressed as $C_i = \frac{|V_i \cap V_{i+1}|}{|V_i|}$. As we expected, there is a high continuity, in most cases larger than $2/3$ – the cases where it is lower can be explained by the passing of some Senators or other unfortunate events. It is also interesting to notice that this score tends to increase as time passes.

We have represented on Figure 7 the cumulative distribution of the number of terms served by each Senator. More than a third of all Senators have served more than five terms and almost half of the Senators have served at least three terms. Contrast this with the fact that a U.S. President cannot be elected for more than two terms since the passage of the Twenty-second amendment. The notion that there should be a term limit in Congress was brought forth by the Republican Party in the 1990s but the proposal fell through in the House.

We shall now quantify the evolution of the communities in two different ways. First, similar to the way we have defined continuity for the whole Senates, we shall define a metric of continuity between to communities of two consecutive Congresses. Let us consider the communities $(S_{i,j})$ and $(S_{i+1,j})$ of the i^{th} and $i+1^{\text{th}}$ Congresses. We shall define then continuity between two communities $S_{i,j}$ and $S_{i+1,k}$ as $c(S_{i,j}, S_{i+1,k}) = \frac{|S_{i,j} \cap S_{i+1,k}|}{\left|S_{i,j} \cap \bigcup_l S_{i+1,l}\right|}$. That is, the ratio of Senators present in both groups compared to those present in the oldest one and which are also active in the second Congress. The idea is to compare the dispersion of the Senators present in a given group, this is why we restrict ourselves to those who are present in both Congresses i and $i+1$. For each community S_1, we shall say that is *successor* is the community S_2 for which $c(S_1, S_2)$ is maximal. Figure 8 displays the cumulative distribution of community continuities between each group and its successor. It is particularly notable that in more than 90% of cases, half of the members present in a community are also present in its successor – once again, only counting those present in both sessions.

Another way of looking at this question is at the level of the Senators themselves. We shall say that two Senators are *co-present* to a certain degree if they belong to at least one same community. Let u and v be two senators, we define their *co-presence* as the number of times they appear in the same community divided by the number of times they are active in a same Congress. Given that a Senator might be in several different communities, we shall count one presence for each community, and thus a Senator can be virtually present more than once during one given Congress. More formally, let $(S_{i,j})$ be the set of communities for the graph G_i, we write $\mathcal{T}_{u,v}$ the number of times u and v appear in a same community. We similarly define $\mathcal{S}_{u,v}$, the number of times u and v appear in the same session. We can then write the co-presence of u and v as: $P(u, v) = \frac{\mathcal{T}_{u,v}}{\mathcal{S}_{u,v}}$

Figure 9 represents the cumulative distribution of the value of the co-presence for a selected subset of pairs of Senators. We have voluntarily excluded the pairs who never appear in the same Congress, that is $\mathcal{S}_{u,v} = 0$ as it would make no sense to compare their communities. Next, we have also removed the pairs which are never in the same community ($\mathcal{T}_{u,v} = 0$), as their inclusion bring no information on the stability of the communities. Finally we have chosen to exclude the pairs who only appear in one Congress and belong to the same community ($\mathcal{T}_{u,v} = \mathcal{S}_{u,v} = 1$), although their inclusion would artificially increase the cumulative distribution we believe it would provide no insight on the actual dynamic aspects of communities given that the information is only extracted from one same timeslice. Their remains the pairs of Senators who appear in at least two Congresses and who are at least once in the same community. We observe that more than 30% of pairs of Senators are stable through time in respect to their communities, that is to say that they have a co-presence of 1 and therefore always appear together in the same community. Moreover, 75% of the pairs have a co-presence greater than 0.5, meaning that 75% of pairs of senators who appear together in at least one community are in the same community in the majority of the Congresses they are active in.

We have described the evolution of the number of communities of the agreement graphs through time, which we explained by refering to the history of the United States Political system and we have exhibited the continuity in Senate membership between Congresses. We have then shown that the communities which were found using C^3 present a certain stability, as they present a high continuity and that Senators appearing together in one community tend to be in the same communities during other Congresses. It is most notable to observe this kind of stability given that the data analysis done on each graph was made independently from the other graphs, which leads us to validate the use of C^3 to compute the communities of agreement.

6 The Blurry Line Between Parties

Until now, we have justified the number of communities by refering to political parties. Fortunately, we have access to the political affiliation of each Senator in our graph which means we can validate that intuition. We shall say that a party is *dominant* in a community if it has the largest representation, and we shall call the *domination ratio* of a community the quotient of the number of members of the dominant party in the community divided by the size of the community. On Figure 10 we have represented, for each session, the average of the domination ratio over all communities. The majority of communities have an average domination ratio of 70%, which means that in most cases one can identify the community to the political party.

In particular, let us look more precisely at a subset of the data, ranging from the 105th Congress (1997–1999) to the 112th Congress (2011–2013). The sizes of the communities as well as the number members of each party represented in the community are given in Table. 2 and a visual representation of those communities are given in Figure 11. First notice that although the communities are allowed to overlap, there are only five cases where we witness an overlap, three of which being because one individual is part of the two communities, one because seven are shared between the two groups of the 111th Congress (2009–2011) and the largest overlap is attained during the 108th Congress (2003–2005) where 10 Senators are part of both communities.

As stated before, each of those communities has a clearly dominant party, be it the Democrat Party or the Republican Party. It is however interesting to observe three things. First, even though the communities have a large majority belonging to a same party, there are members of the other party which are more in agreement with there opponents than their political family, and this even in cases where there is no overlap between communities. For example, in the 106^{th} and 107^{th} Congresses, the democrat which is in the Republican community is Zell Miller, who has frequently criticized the Democratic Party since 2003, backed the Republican President over the Democratic nominee in the 2004 presidential election, has publicly supported several Republican candidates and serves as the national co-chair to the campaign of Republican presidential candidate Newt Gingrich.

During the 108^{th} Congress (2003–2005) there is a large ovelap between the two communities, leading to the presence of 18 Democrats in the Republican agreement community. We have found no satisfactory explanation to this observation. The first line we pursued while trying to understand that result is that the United States invaded Iraq in 2003, but it turns out that more than half of those 18 Democrat Senators were vocally opposed to the use of force to overthrow the Iraqi government and thus would have no reason whatsoever to vote in agreement with the Republicans on those texts. Furthermore, we have found no trace of any particular event which might explain this result and thus the question of what happened in 2003–2004 in the Senate remains open.

Finally, there are some Congresses where there is a large number of Republicans in the otherwise Democrat dominated comunity. For example, in the 110^{th} Congress (2007–2009) there are 11 Republicans which are more in the Democrat community of agreement. A list of facts about 8 of those Senators is given in Appendix E in order to explain their presence in a Democrat community.

We have observed that, although most communities are largely dominated by a party, there are some cases where Senators from other parties are present. By looking into the political profile of those seemingly displaced individuals, we have explain their placement by a disalignment between the Senators and their official affiliation. Note that in some cases we were however unable to find a plausible explanation, such as for example as to why 18 Democrats have been grouped into the Republican community in the 108^{th} Congress. This adds to the validation of C^3 to compute community of agreement, given that no party information had been used to calculate the communities.

Conclusion

In this article we have first proved that the problem of finding communities with maximal cohesion is an $\mathcal{NP}$-hard problem. We have then also proven that the dual problem of minimizing the cohesion at a fixed size is also $\mathcal{NP}$-hard. In order to compute social communities we have presented a heuristic, C^3, which covers a graph by expanding around selected edges and then combines communities depending on an overlap parameter.

Finally, we have applied this heuristic to an agreement graph of U.S. Senators votes in which we have observed that the communities found by C^3 are relatively stable through time despite being computed independently from one graph to the other. We have also shown that the communities can be assimilated to their dominant political party and found an explanation as to the presence of members of one party in a community dominated by another one.

We had previously shown that the cohesion was a good indicator of subjective perception of communities and the two last results lead us to believe that, although being a heuristic, C^3 yields communities which make sense from a social standpoint.

References

[1] Adrien Friggeri, Guillaume Chelius, and Eric Fleury. Triangles to Capture Social Cohesion. In *2011 IEEE Third International Conference on Social Computing (SocialCom 2011)*, pages 258–265, 2011.

[2] John M Lewis and Mihalis Yannakakis. The node-deletion problem for hereditary properties is NP-complete. *Journal of Computer and System Sciences*, 20(2):219–230, 1980.

A Proofs

A.1 Proof of Lemma 1.1

By contradiction, suppose that $\mathcal{C}(S_1) \leq \mathcal{C}(S_1 \cup S_2)$ and $\mathcal{C}(S_2) \leq \mathcal{C}(S_1 \cup S_2)$, which we can rewrite as:

$$\frac{\bigtriangleup(S_1)^2}{\binom{|S_1|}{3}} \leq \Big(\bigtriangleup(S_1) + \wedge(S_1)\Big)\,\mathcal{C}(S_1 \cup S_2) \tag{1}$$

$$\frac{\bigtriangleup(S_2)^2}{\binom{|S_2|}{3}} \leq \Big(\bigtriangleup(S_2) + \wedge(S_2)\Big)\,\mathcal{C}(S_1 \cup S_2) \tag{2}$$

By summing (1) and (2) it comes that:

$$\frac{\bigtriangleup(S_1)^2}{\binom{|S_1|}{3}} + \frac{\bigtriangleup(S_2)^2}{\binom{|S_2|}{3}} \leq \Big(\bigtriangleup(S_1) + \wedge(S_1) + \bigtriangleup(S_2) + \wedge(S_2)\Big)\,\mathcal{C}(S_1 \cup S_2)$$

Now, given that S_1 and S_2 are disconnected, we have:

$$\bigtriangleup(S_1) + \bigtriangleup(S_2) = \bigtriangleup(S_1 \cup S_2)$$
$$\wedge(S_1) + \wedge(S_2) = \wedge(S_1 \cup S_2)$$

Therefore,

$$\begin{aligned}\frac{\bigtriangleup(S_1)^2}{\binom{|S_1|}{3}} + \frac{\bigtriangleup(S_2)^2}{\binom{|S_2|}{3}} &\leq \Big(\bigtriangleup(S_1 \cup S_2) + \wedge(S_1 \cup S_2)\Big)\,\mathcal{C}(S_1 \cup S_2)\\ &\leq \frac{(\bigtriangleup(S_1) + \bigtriangleup(S_2))^2}{\binom{|S_1|+|S_2|}{3}}\end{aligned}$$

Furthermore, given that $|S_1|, |S_2| > 1$, the following holds:

$$\binom{|S_1|}{3} + \binom{|S_2|}{3} < \binom{|S_1| + |S_2|}{3}$$

From there it comes:

$$\frac{\bigtriangleup(S_1)^2}{\binom{|S_1|}{3}} + \frac{\bigtriangleup(S_2)^2}{\binom{|S_2|}{3}} < \frac{(\bigtriangleup(S_1) + \bigtriangleup(S_2))^2}{\binom{|S_1|}{3} + \binom{|S_2|}{3}}$$

Which simplifies to:

$$\left(\binom{|S_2|}{3}\bigtriangleup(S_1) - \binom{|S_1|}{3}\bigtriangleup(S_2)\right)^2 < 0$$

Hence the contradiction. Therefore, for all $S_1, S_2 \subseteq V(G)$, disconnected:

$$\mathcal{C}(S_1) \leq \mathcal{C}(S_1 \cup S_2) \Rightarrow \mathcal{C}(S_2) > \mathcal{C}(S_1 \cup S_2)$$

A.2 Proof of Theorem 1.3

Let $K \subseteq V$, be a clique of size $|K| = k$ in G. Given that no node or edge are deleted when constructing G', G is a subgraph of G' and thus K is a clique in G' and $\bigcirc\!\!\!\!\!\triangle_{G'}(K) = \binom{k}{3}$.

Moreover, by construction, $G'[V]$ is a clique and for all u in K, the neighbors of u are also in V. Therefore, each edge in K forms one triangle with each vertex in $V \setminus K$, which leads to $\underline{\triangle}_{G'}(K) = \binom{k}{2}(n-k)$. Finally, this gives a cohesion:

$$\mathcal{C}_{G'}(K) = \frac{\binom{k}{3}}{\binom{k}{3} + \binom{k}{2}(n-k)}$$

Conversely, let $S \subseteq V'$ be a connected set of vertices such that $\mathcal{C}_{G'}(S) \geq \frac{\binom{k}{3}}{\binom{k}{3}+\binom{k}{2}(n-k)}$. We will show that S is a clique of size larger than k and that $S \subseteq V$. First note that $|S| \geq 3$, because by definition, if $|S| < 3$, $\mathcal{C}_{G'}(S) = 0$ which would lead to a contradiction.

First, suppose that S is not a clique in G, then let us distinguish two cases:

1. If $S \subseteq V$ and S is not a clique, then S contains two vertices $u, v \in V^2$ such that $uv \notin E$.

2. If $S \subsetneq V$, then $\exists u \in S \setminus V$, and S being connected, there exists $v \in V'$ such that $uv \notin E$.

Therefore, if S is not a clique in G, it contains an edge $uv \notin E$ and by construction, this edge belongs to at least $2\binom{n}{3}^4$ triangles, which leads to:

$$\begin{aligned}
\bigcirc\!\!\!\!\!\triangle_{G'}(S) + \underline{\triangle}_{G'}(S) &\geq 2\binom{n}{3}^4 \\
\mathcal{C}_{G'}(S) &\leq \frac{\bigcirc\!\!\!\!\!\triangle_{G'}(S)^2}{2\binom{|S|}{3}\binom{n}{3}^4} \\
&\leq \frac{1}{2\binom{n}{3}^2} \\
&< \frac{\binom{k}{3}}{\binom{k}{3} + \binom{k}{2}(n-k)}
\end{aligned}$$

Hence the contradiction, therefore S must be a clique in G. From there it comes that:

$$\mathcal{C}_{G'}(S) = \frac{\binom{k'}{3}}{\binom{k'}{3} + \binom{k'}{2}(n-k')}$$

where $k' = |S|$. Therefore:

$$\begin{aligned}
\mathcal{C}_{G'}(S) \geq \frac{\binom{k}{3}}{\binom{k}{3} + \binom{k}{2}(n-k)} &\Leftrightarrow \frac{\binom{k'}{2}(n-k')}{\binom{k'}{3}} \leq \frac{\binom{k}{2}(n-k)}{\binom{k}{3}} \\
&\Leftrightarrow \frac{n-k'}{k'-2} \leq \frac{n-k}{k-2} \\
&\Leftrightarrow k' \geq k
\end{aligned}$$

Therefore, we can now conclude that if there exists a connected set S in G' with cohesion $\mathcal{C}_{G'}(S) \geq \frac{\binom{k}{3}}{\binom{k}{3}+\binom{k}{2}(n-k)}$, then S is a clique of size at least k in G, and thus there exists a clique $K \subseteq S$ of size k in G.

B Algorithms

B.1 Transform an instance of Clique into an instance of Connected-Cohesive

```
Require: $G = (V, E), k \in \mathbb{N}$
  $W := \emptyset$
  $E' := E$
  for $uv \in V^2 \setminus E$ do
      let $K$ be a clique of size $2\binom{n}{3}^4$
      $W \leftarrow W \cup K$
      $E' \leftarrow E' \cup \{uv\} \cup (\{u, v\} \times K)$
  return $G' = (V \cup W, E'), \lambda = \frac{\binom{k}{3}}{\binom{k}{3}+\binom{k}{2}(n-k)}$
```

B.2 Valid Candidates

```
function isCandidate($S \subseteq V, u \in V$)
    return false if $u \notin \mathcal{N}(S)$
    return false if $u \in S$
    return true if $\textcircled{\vartriangle}(S) = 0$ and $\overset{\vartriangle}{\ominus}(S) < \overset{\vartriangle}{\ominus}(S \cup \{u\})$
    return false if $\textcircled{\vartriangle}(S) = \textcircled{\vartriangle}(S \cup \{u\})$
    return true if $\mathcal{C}(S \cup \{u\}) \geq \mathcal{C}(S)$
    return true if $\textcircled{\vartriangle}(S \cup \{u\}) \geq \frac{|S|+1}{|S|-2} \textcircled{\vartriangle}(S)$
    return false
```

B.3 Communitize

```
function Communitize($S \subseteq V$)
    $B \leftarrow S$
    $C \leftarrow \{ u \in V \mid$ isCandidate$(S, u) \}$
    for all $u \in C$ sorted as defined above do
```

```
        if $u$ is not marked as visited then
            mark $u$ as visited
            $B' \leftarrow$ COMMUNITIZE$(S \cup \{u\})$
            if $\mathcal{C}(B) \leq \mathcal{C}(B')$ then
                $B \leftarrow B'$
    return $B$
```

B.4 Update the Cohesion in place

```
function UPDATESET$(S \subseteq V, u \in V, \delta)$
    $🛆 \leftarrow 🛆 + \delta\, 🛆_u$
    $⏏ \leftarrow ⏏ + \delta \left( ⏏_u - 🛆_u \right)$
    for all $v \in \mathcal{N}(u)$ do
        for all $w \in \mathcal{N}(u) \cap \mathcal{N}(v)$ do
            if $v \in S$ then
                $🛆_w \leftarrow 🛆_w + \delta$
                $⏏_w \leftarrow ⏏_w - \delta$
            else
                $⏏_w \leftarrow ⏏_w + \delta$
```

B.5 Cover

```
function COVER$(S \subseteq V)$
    $C \leftarrow \emptyset$
    $M \leftarrow \emptyset$
    for all $u \in S \setminus M$ in increasing order of $d(u)$ do
        $C_u \leftarrow \emptyset$
        for all $v \in \mathcal{N}(u) \setminus M$ do
            $M \leftarrow M \cup$ COMMUNITIZE$(\{u, v\})$
        if $C_u = \emptyset$ then
            for all $v \in \mathcal{N}(u) \cap M$ do
                $M \leftarrow M \cup$ COMMUNITIZE$(\{u, v\})$
        $c \leftarrow$ element of $C_u$ with maximum cohesion
        $C \leftarrow C \cup \{c\}$
    return $C$
```

C Figures

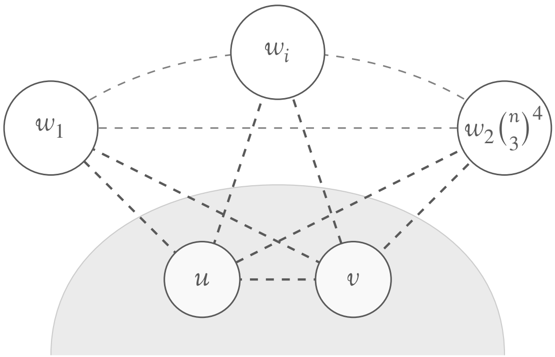

Figure 1: Illustration of Algorithm B.1. At this step, we join u and v, add a clique of size $2\binom{n}{3}^4$ to the network, and join u and v to all vertices in the added clique.

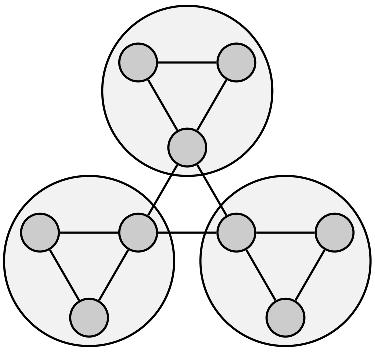

Figure 2: Those four cliques would not be found if only expanding aroung one node, only the three encircled ones would.

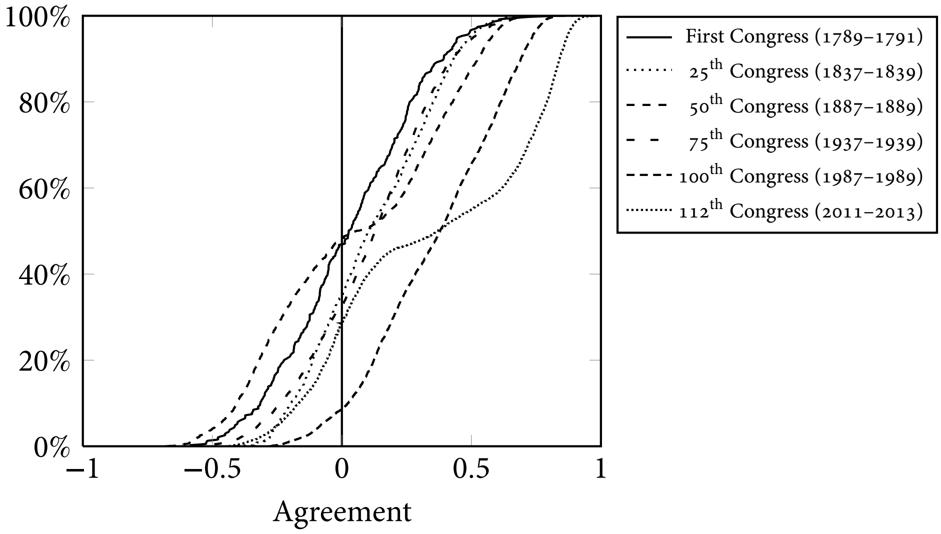

Figure 3: Cumulative distribution of the weights in Senate agreement graph for six different Congresses.

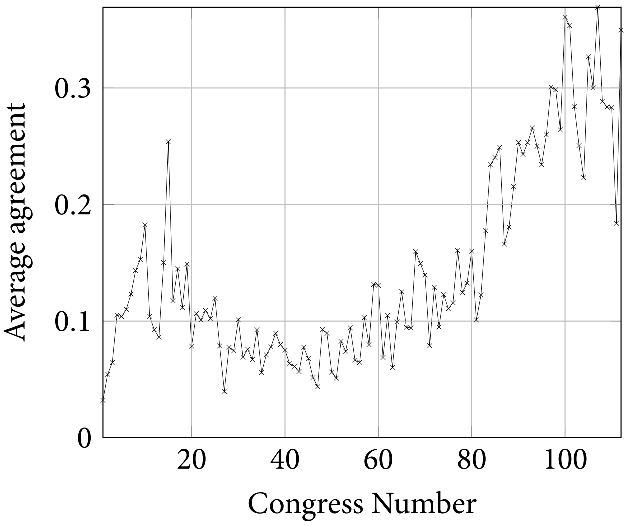

Figure 4: Evolution through time of the average agreement weight between Senators.

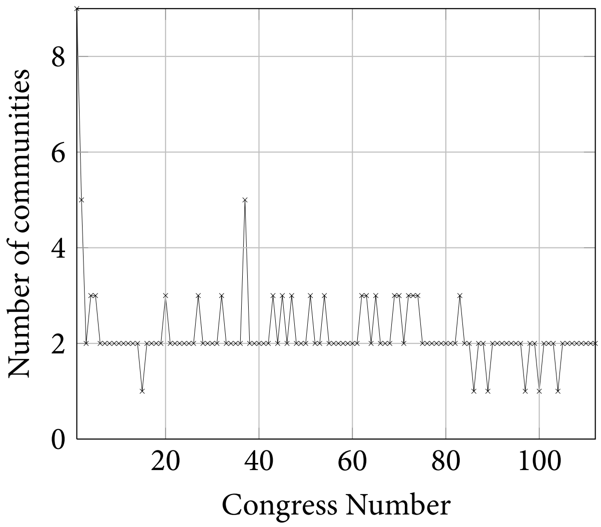

Figure 5: Evolution through time of the number of communities in the Senate agreement graphs.

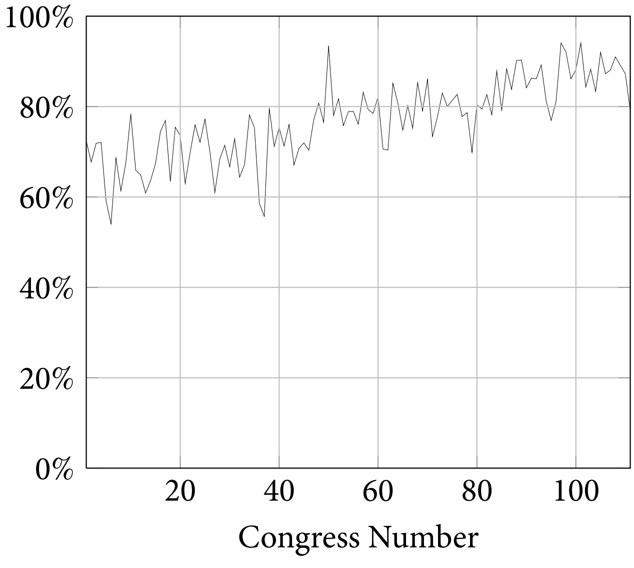

Figure 6: Evolution through time of the proportion of Senators remaining in office between two Congresses.

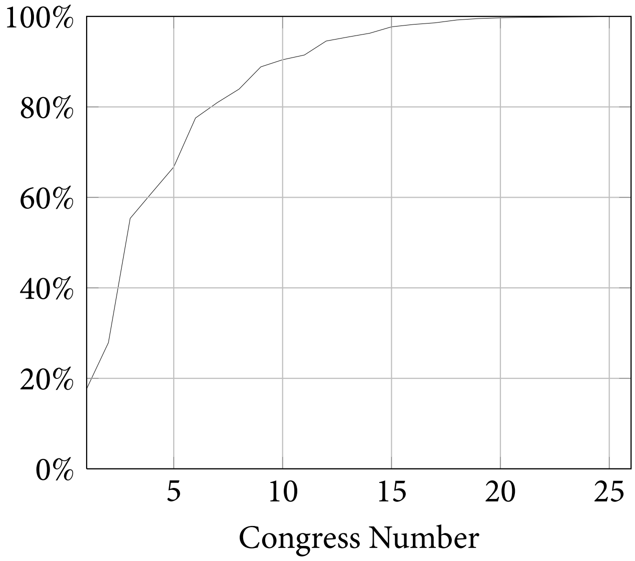

Figure 7: Cumulative distribution of the number of terms for each Senator.

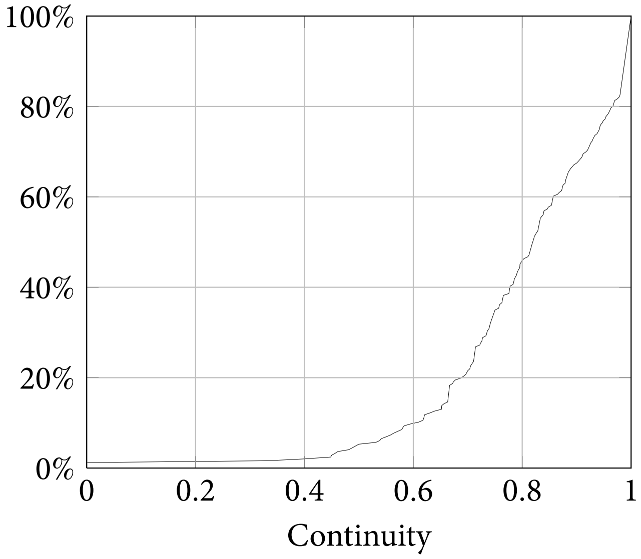

Figure 8: Cumulative distribution of the values for community continuity between a community and its successor.

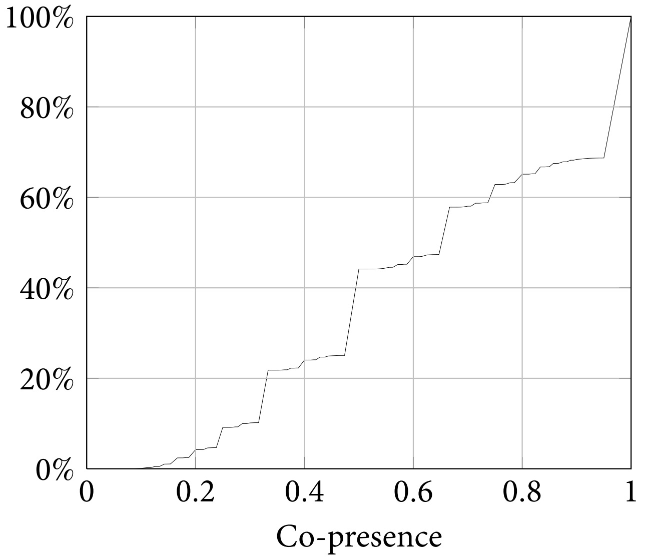

Figure 9: Cumulative distribution of the co-presence ratio for pairs of Senators present in at least one community together.

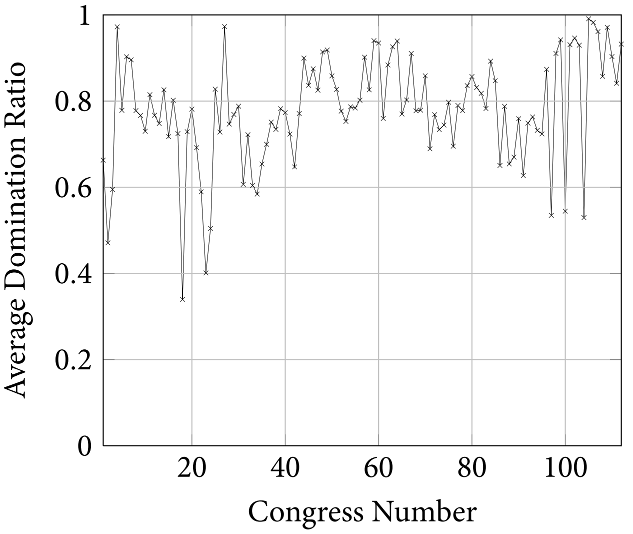

Figure 10: Evolution through time of the average domination of the communities by one political party.

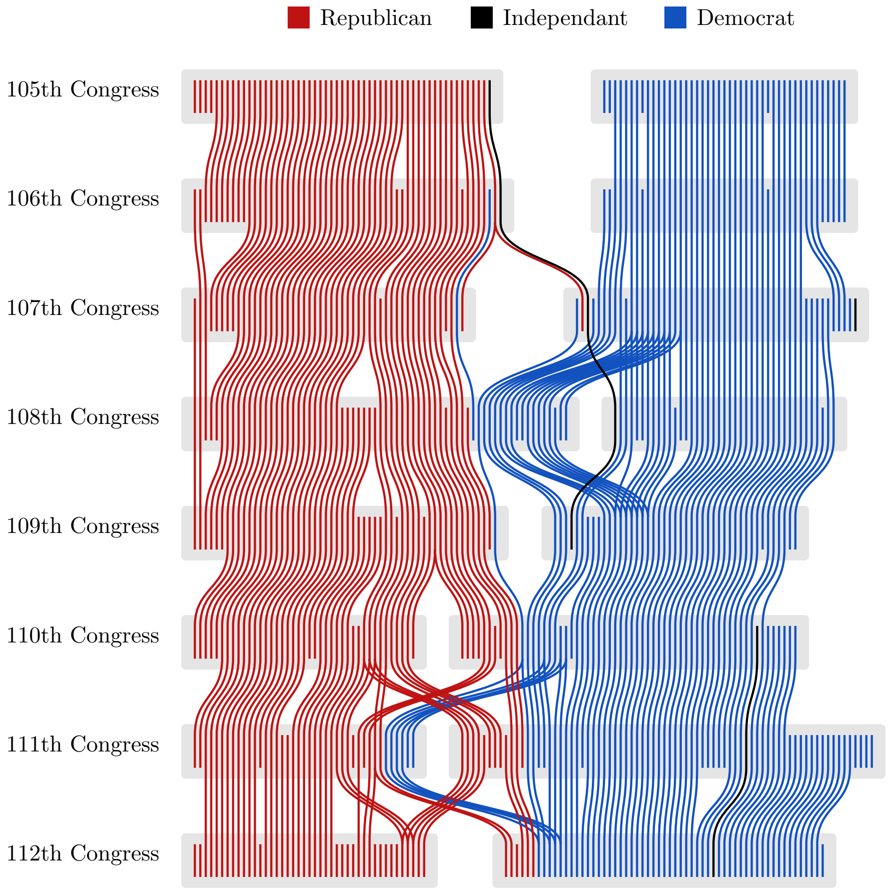

Figure 11: Political party breakdown of the communities of the Senate agreement graph for the 8 last Congresses. Each path represent a Senator and the color indicates the Political party they belong to. Each box is a community and each row of boxes represent a particular Congress. Notice that the overlaps between communities are visible here (*e.g.* where a Senator path forks into two different communities).

D Tables

	$\bigcirc\mkern-14.5mu\triangle < 0$	$\bigcirc\mkern-14.5mu\triangle \geq 0$
$\underline{\triangle} < 0$	$\frac{\bigcirc\mkern-14.5mu\triangle}{\binom{n}{3}} \frac{\bigcirc\mkern-14.5mu\triangle}{\bigcirc\mkern-14.5mu\triangle + \underline{\triangle}} \leq 0$	$\frac{\bigcirc\mkern-14.5mu\triangle}{\binom{n}{3}} \geq 0$
$\underline{\triangle} \geq 0$	$\frac{\bigcirc\mkern-14.5mu\triangle}{\binom{n}{3}} \leq 0$	$\frac{\bigcirc\mkern-14.5mu\triangle}{\binom{n}{3}} \frac{\bigcirc\mkern-14.5mu\triangle}{\bigcirc\mkern-14.5mu\triangle + \underline{\triangle}} \geq 0$

Table 1: Impact of signed triangles weights on the Cohesion.

105th Congress (1997–1999)	100 members	
	45 Democrat	55 Republican 1 Independant
106th Congress (1999–2001)	102 members	
	45 Democrat	56 Republican 1 Independant 1 Democrat
107th Congress (2001–2003)	101 members	
	49 Democrat 2 Independant 1 Republican	49 Republican 1 Democrat
108th Congress (2003–2005)	100 members	
	40 Democrat 1 Independant	51 Republican 18 Democrat
109th Congress (2005–2007)	101 members	
	44 Democrat 1 Independant	55 Republican 1 Democrat
110th Congress (2007–2009)	102 members	
	50 Democrat 1 Independant 11 Republican	41 Republican
111th Congress (2009–2011)	110 members	
	63 Democrat 1 Independant 12 Republican	35 Republican 6 Democrat
112th Congress (2011–2013)	101 members	
	51 Democrat 1 Independant 6 Republican	43 Republican

Table 2: Political party breakdown of the communities of the Senate agreement graph for the 8 last Congresses.

E Profiles of 8 Republicans in a Democrat community

Charles Hagel was critic of the Bush Administration among other things over the Iraq War, which in 2005 he compared to Vietnam. Furthermore, he has repeatedly taken issue with the Bush Administration while in office up to the point of rating, in November 2007, the Bush administration "the lowest in capacity, in capability, in policy, in consensus almost every are" of any presidency in the last forty years and adding in 2008 "I have to say this is one of the most arrogant, incompetent administrations I've ever seen or ever read about.". During the 2008 Presidential Election, he also revealed he was open to running as Vice-President with the Democrat nominee.

George Voinovich also had issues with the Bush Administration, in particular in terms on international politics and the Iraq War. In 2007 he did not share President George W. Bush's optimism about the effectiveness of sending more troops to Iraq. In 2008 he said at a hearing before the Senate Foreign Relations Committee regarding the war in Iraq: "We've kind of bankrupted this country" through war spending. "We're in a recession...and God knows how long it's going to last.". Finally, Voinovich has been know to oppose lowering taxes and frequently joined the Democrats on tax issues.

John Warner is a moderate Republic and has centrist stances on many issues, to the point that he faced opposition of other members of its own party when he decided to run for re-election for a fourth term in the Senate in 1996 – he was a Senator from 1979 to 2009.

Olympia Snowe and **Susan Collins** are the two Republican Senators from Main which are both reagarer to be leading moderates within their party, and were described in 2009 as "nearly the last survivors of a once common species of moderate Northeastern Republican".

John McCain was also described as a moderate although he began adopting more orthodox conservative views since his loss in the 2008 Presidential Elecion – some analysis point his centre-right image as one of the reasons he lost. Before 2008, it was said that while McCain usually tended towards conservative positions, he was not "anchored by the philosophical tenets of modern American conservatism". He was also part of a group of senators which came to be known as the Gang of 14 – of which Olympia Snowe, Susan Collins and John Warner were also a part of – which had brokered a compromise in difficult times between Republicans and Democrats during the previous Congress.

Gordon Smith is often described as politically moderate, although he also has strong conservative credentials. He was placed in the exact ideological center of the Senate by a 2006 National Journal congressional rating and he started criticizing the Iraq War in 2006, after supporting it for four years, going as far as saying that then current policy in Iraq "may even be criminal".

Norm Coleman had been a Democrat until he switched parties in December 1996, although still being considered one of the most liberal Republican in the Senate. While running for mayor in 1993, he wrote in a letter to the City

Convention Delegates in which he says: “I am a lifelong Democrat. [...] my commitment to the great values of our party has remained solid.”.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399