Article

Detecting opportunities to moderate the Social Media debate with Adiabatic Quantum Computing: a COVID-19 case

Juan Bernabé-Moreno [1,2,3]

[1] University of Oxford;
[2] University of Granada;
[3] E.ON SE

Abstract: Social media has become the most influential battleground for political and societal debate. The reach, the speed, and the fact that everybody can participate make this medium very popular but also highly manipulative, subject to biases and extremely precious to control. Misinformation spreads in an unforeseeable way. Sometimes it just follows usual non-malicious social interaction patterns, but often a targeted and designed-for-purpose manner. In this paper, we propose a novel method to identify the most effective opportunities to moderate the social media debate at a large scale using the properties of social signed graphs, such as structural balance. After building the topic universe for a particular political debate in a given time frame, we extract the most active and influential users and mine their position regarding the topic based on their interactions with other users. The result is a signed graph where each edge represents either support or challenge between a pair of users in the context of the topic. This graph can be modelled as an Ising model, with the goal of minimizing its energy state by using the advantages of Adiabatic Quantum Computing. Applying this principle, we use a Quantum Annealing implementation to solve the problem of finding the edges violating the structural balance of this graph, which in classic computing is known to be an NP-hard problem. We then provide a non-intrusive moderation technique for those violations to mitigate or increase the structural imbalance, depending on the desired strategy. The methodology is shown in the context of the COVID-19 skeptics' debate in Twitter and implemented using a D-Wave 2000Q-6 quantum annealer.

1. Introduction

In the most recent years and more intensively than ever before, political and societal issues are debated in the social media (SM) networks. Everybody with an account can not only follow the debate but also take an active role and shape it. Unlike other more traditional media, the SM networks are designed to be open and democratic (everybody has the opportunity to share their own opinions and views). The importance or weight of these opinions depends on the properties of the author's subnet itself -the number of connections, the importance of the users to which it is connected-, hir or her level of activity, and the very nature of the social network, etc [1]. Corporations of all sizes, governmental institutions, NGOs, academia, etc. are intensifying their presence and activity in SM platforms to increase brand awareness, improve loyalty or position themselves as thought-leaders. Managing SM has become an integral part of the corporate identity and the marketing strategy of any modern company or institution [2,3].

The former US President Donald Trump has been using his personal Twitter account not only as the medium to broadcast his official views but also to take official actions (e.g.: nominating the FBI director, announcing his administration's transgender military ban, etc). Hence, the National Archives concluded that President Trump's tweets are "official records that must be preserved under the Presidential Records Act.", which has been the subject of countless discussions [4]. In one way or another, it has catapulted the importance of SM platforms -especially Twitter and Facebook-, in the political and societal debate.

We are witnessing the proliferation of more or less evident SM manipulative attempts of all kinds [5,6], some of them with probably history-changing consequences, such as the *Cambridge Analytica* scandal [7,8]. It has been pushing the debate on how to balance access to free speech vs. the uncontrolled spread of fake information with manipulating purposes. Despite the efforts of many countries to pass laws to discourage the production and sharing of fake news (e.g.: the "Protection from Online Falsehoods and Manipulation Act" in Singapore -April 2018-, "Fake News Law" in France -November 2018- or the "Network Enforcement Act" in Germany -June 2017-, etc), substantial problems remain. Among the most pressing ones, we encounter detecting fake news and attempts of manipulation, response time for enforcement -as misinformation spread happens real-time-, identifying the real authors, shutting down bots, fighting crowd-turfing [9,10].

In addition to the aforementioned counter-measures, more and more parties see themselves obliged to engage. Yet, the active participation in the SM debate, with potentially millions of participants, is time-consuming and without a proper strategy, any kind of effort is not likely to pay off or have any major impact.

Handling misinformation requires near real-time monitoring and consciously selecting the most effective interventions in an overly dynamic environment, introducing the need for constantly re-evaluating the moderation strategy. The complexity increases when, in addition to the ones actively participating (liking, sharing, commenting, etc), we increase the scope of our strategies to take into account all the followers who are exposed to the streaming of interactions related to the debate [11].

Tailoring an optimized engagement strategy requires first identifying which topics are part of the debate and to which extent. Some debates emerge around a few well-defined hashtags and their variations, others are more scattered and quite often, several debates tend to overlap and become difficult to disentangle. Participants are creating posts, interacting with the posts others created, mentioned in posts, etc in the context of one or many of the identified topics [12,13].

Secondly, we need to understand the positioning of any of the participants towards the topic (in favour, neutral, against) as well as their ability to influence others. The underlying structural balance of the SM network involving all users and their interactions related to a topic can be employed to develop holistic moderation strategies [14,15]. Once we have the most critical users mapped, it is important to identify in which cases the debate is being neutralized by other users (is *balanced*), and if it is imbalanced, in which sense. The engagement strategy shall then prioritize addressing the violations of the balanced state [16] and be implemented in the most non-intrusive way possible (as it might be perceived by the users as censorship [17]).

In this paper we proposed a novel holistic method to create an effective strategy to moderate the SM debate, consisting of five essential parts:

1. **Topic modelling:** given the context of a social debate, to identify what are the leading topics discussed in this debate.
2. **Topic membership definition:** given a SM interaction $it_i \in I$ and a particular topic $T_j \in T$, to determine if, and to which degree, it_i belongs to the particular topic T_j.
3. **Opinion polarity graph construction:** given the sub-network of all users who created or engaged with any interaction related to the topic T_j over a particular period, the procedure to extract a signed graph evaluating the *polarity Engagement* of every pair of users and applying *polarity disambiguation* if more than one interaction is registered.
4. **Quantum-based structural balance violations extraction:** to extract the set of interactions which violates the structural balanced of the *Opinion polarity graph* -which given the NP-hard nature of the problem, we model it as an *Ising system* and run it on a quantum annealer [18]-.

5. **Non-intrusive moderation:** having a set of structural balance violations and knowledge about the network, to create third party interactions between the involved users with the aim of triggering positive reactions and therefore alleviating the structural imbalance of the *Opinion polarity graph*.

In addition to the originality of the methodology just described, this paper's contribution encompasses several novel aspects: a) we have introduced several concepts, such as *polarity engagement*, *polarity disambiguation* in the context of SM, b) we have designed an algorithm to build the *Opinion polarity graph* from a set of interactions and users in a SM network, c) we have employed a Quantum annealing approach to solving a moderation problem in a topic-specific SM network and finally, we have suggested a non-intrusive moderation strategy, which minimizes the number of interactions required to restore the SM structural balance.

Our work is structured as follows: we first discuss the rationale of our attempt and explain the novelty of our approach in the introduction, to then share the background information supporting our research in section 2. Following that, in section 3, we formally describe our method, providing some preliminary definitions and showing how to define our *Opinion polarity graph* and solve the structural balance problem using Quantum annealing technology to find the interactions requiring moderation. We then present the design of the system and implement our approach in section 4 and discuss the results obtained for the Twitter debate around the COVID-19 topic in section 5. We conclude by providing our concluding remarks and pointing to further research lines.

2. Background

2.1. On Social Media manipulation and moderation

SM has been the target of systemic manipulation as controlling the opinions of the digital masses is commonly seen as a new form of power [6]. The relevance of this matter has triggered a very prolific research activity to better understand implications, develop identification strategies, implement countermeasures, etc. In [19], for instance, the authors warn about the challenges that a digital democracy and the technological optimism brought by the open and collaborative networking characteristics of SM.

Yang et al. discuss in [20] how SM play an increasingly important role in political communication and how private individuals but also politicians share their views about politics and push to mobilize and protest against social issues. A thorough analysis based on the agenda-setting theory [21] using Twitter data reveals the patterns of daily political discussion, uncovering the main topics of attention and interest of these actors.

Manipulations are usually organized and often times have established groups, such a government, military or political party teams committed to manipulating public opinion over social media. The Oxford Internet Institute has been cataloguing what they called *cyber troops manipulations* since 2017 [5].

Manipulative attempts through targeted spread of misinformation involves three major tasks: detection of fake or manipulative content and its authorship, understanding and tracing of the propagation mechanism and design and execution of a moderation plan.

Automatic detection methods have been developed to identify fake news in social media. In [22], the authors compare and benchmark different machine learning approaches, encompassing dictionary-based and several deep learning methods. Gillespie et al. [23] discuss the need for Artificial Intelligence approaches to reach scale in the moderation but also provide some arguments against automatic content moderation. In [24] the authors present a method to incorporate human supervision, where human content moderators are aided by a supervised machine learning system which provides suggestions regarding the relevance and categorization of the content. Yadin and his co-authors [25] suggest the use of crowd-sourcing in combination with NLP algorithms to compile and revise content moderation guidelines semi-automatically.

Moderating manipulative attempts, in addition to the underlying complexity, pose several additional challenges. In [26], the authors analyse the reactions of the users to intrusive moderation, such as censoring, suspension or shadow-banning. In [17], Baker et al. discussed the key challenges related to the harm-based moderation policies to address the misinformation spread problem, covering also the COVID-19 case.

In this work, we focus on a non-intrusive moderation method to prevent the perception of centralized censorship and rather activate other users to counter-fight misinformation.

2.2. On Signed Social Networks and structural balance

The first large scale approach to study the interplay between negative and positive relationships and their impact on the structure of SM networks has been proposed by Leskovec et al [27]. The authors connect SM to theories of signed networks from social psychology and also tackle the evolution of the ties.

The network structural balance is a well-known and highly studied property of the signed networks. In social networks (and by extension, in SM networks), structural balance, also called social balance, is formulated to understand the stability or tensions in population systems [28,29]

In [30], the author explores applications of signed network analysis methods that exploit structural balance, such as finding communities, drawing signed social networks, and solving the problem of link prediction.

The approach to analyse the structural balance of the graph focuses on identifying how many and which edges should change the sign or be pruned to restore the balance in the network. We call structural conflicts of the signed network, the minimal numbers of edges, which should be deleted to make the network balanced. While F. Heider in 1956 in his seminal work *Attitudes and cognitive organization* [31] provided a method to assess if a signed network is structural-balanced or not, detecting the structural conflicts is an NP-hard problem from the computational perspective [32]. Similarly complex is the computing of the equivalent *frustration index*, which indicates the minimum number of edges whose removal (or equivalently, negation) results in balance restoration.

2.3. On Signed Graphs Structural Balanced, Ising systems and Quantum annealing

There is a close relationship between the frustration index of signed graphs, are the ground-state properties of Ising models [33]. The key objective of an Ising model, representing the patterns of atomic magnets based on positive and negative interactions among spins and their nearest neighbours, is finding the spin configurations with the minimum energy, as explained in [34].

Each spin is connected to its neighbours in a grid. Any pair of connected spins can have either an aligned/matched or unaligned/unmatched coupling $J_{ij} = +1$ $(J_{ij} = -1)$ Each spin can either take an upward or a downward configuration. The target is to find a spin configuration for a given set of fixed coupling constants that minimizes an energy function [34]. Frustration manifests in 2 cases: when a matched coupling has a different configuration than the spins configurations of the endpoints or when a mismatched coupling has the same configuration as the spins configurations of the endpoints.

The energy of a spin configuration is calculated based on the Hamiltonian function $H = \sum_i^j J_{ij} s_i s_j$, with $s_1, s_2, ..., s_n$ taking +1 or -1 representing the upward or downward spins, which corresponds to the energy function described in [35] or [36].

As aforementioned, minimizing the H overall potential spin configurations is a NP-hard problem, and exactly where Quantum Computing approaches bring an advantage.

We can define *Quantum Annealing* as a combinatorial optimization technique that leverages tunneling, entanglement, and further quantum mechanical effects to minimize energy-based models.

In [18], the authors present an approach that makes use of quantum annealing hardware for machine learning purposes. At zero temperature the quantum annealing hardware can be used as a heuristic minimizer for Ising energy functions, while at finite temperature, it allows for Boltzmann distribution sampling. The authors demonstrated how the use of quantum mechanical processes outperformed similar approaches based on software simulation to solve these tasks.

Quantum Annealing implementation helps to tackle the NP-hard nature of the structural balance problem and therefore guarantees the feasibility of the approach presented in this paper.

3. Detection of moderation opportunities with quantum annealing

In this section, we first formally introduce a set of concepts and definitions required to model the moderation detection problem. Then we provide a modular view of the components required to implement the detection algorithm as well as the process governing these components.

3.1. Preliminary concepts

Let $G = (V, E)$ be a graph with vertex set V and edge set E such that $n = |V|$ (number of vertices) and $m = |E|$ (number of edges).

We consider **undirected signed networks** $G = (V, E, \sigma)$. The set E of edges can be partitioned into the set of positive edges E^+ and the set of negative edges E^- with $|E^+| = m^+, |E^-| = m^-$,and $|E| = m$ where $m = m^+ + m^-$. The sign function is denoted by $\sigma : E \to \{-1, +1\}$. We represent the m undirected edges in G as ordered pairs of vertices $E = \{e_1, e_2, ..., e_m\} \subseteq \{(i, j)|i, j \in V, i \neq j\}$, where a single edge between nodes i and $j, i \neq j$, is denoted by $(i, j), i \neq j$. We denote the graph density by $\rho = 2m/(n(n-1))$.

$$a_{ij} = \begin{cases} \sigma(i,j) & \text{if}(i,j) \in E \\ \sigma(j,i) & \text{if}(j,i) \in E \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

A walk of length k in G is a sequence of nodes $v_0, v_1, ..., v_{k-1}, v_k$ such that for each $i = 1, 2, ..., k$ there is an edge from v_{i-1} to v_i. If $v_0 = v_k$, the sequence is a closed walk of length k. If the nodes in a closed walk are distinct except for the endpoints, the walk is a cycle of length k. The sign of a walk or cycle is the product of the signs of its edges. Cycles with positive (negative) signs are balanced (unbalanced) [32]. A **balanced graph** is one with no unbalanced cycles.

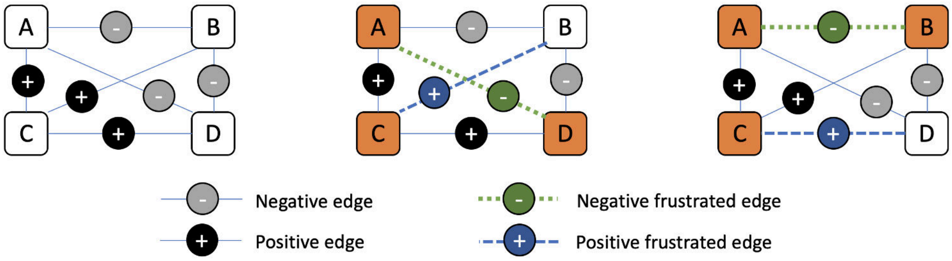


Figure 1. Frustrated graph

For any signed graph $G = (V, E, \sigma)$, we can partition V into two sets, denoted $X \subseteq V$ and $\overline{X} = V \setminus X$. We call X a colouring set and we think of this partitioning as specifying a colouring of the nodes, where 1 demonstrates an example signed graph in which positive and negative edges are represented by solid lines and dotted lines respectively. In addition, we can see two node colourings and the resulting frustrated edges represented by thick lines.

We considered an edge to be satisfied when it is positive and both endpoints have the same value, or it is negative and the endpoints have opposite values. When an edge is not satisfied, we call it frustrated.

We define the **frustration count** $f_G(X)$ as the number of frustrated edges of G under $X : f_G(X) = \sum_{(i,j)\in E} f_{i,j}(X)$ being $f_{i,j}(X)$ is the frustration state of edge (i,j), given by:

$$f_{ij}(X) = \begin{cases} 0 & \text{if } x_i = x_j \; and \; (i,j) \in E^+ \\ 1 & \text{if } x_i = x_j \; and \; (i,j) \in E^- \\ 0 & \text{if } x_i \neq x_j \; and \; (i,j) \in E^- \\ 1 & \text{if } x_i \neq x_j \; and \; (i,j) \in E^+ \end{cases} \tag{2}$$

The **frustration index** $L(G)$ of a graph G can be obtained by finding a subset $X^* \subseteq V$ of G that minimises the frustration count $f_G(X)$ by solving $L(G) = min_{f_{G(X)}} X \subseteq V$. In other words, the frustration index is the smallest number of frustrated edges over all states.

Computing the frustration index is an NP-hard problem, as the frustration index of an all-negative signed graph is equivalent to the maximum cut problem in graph theory, which is NP-hard. The frustration index corresponds to the smallest number of edges whose negation (or deletion, according to the Harary theorem [37]) makes the graph G balanced.

The Ising model can be seen as a signed graph where each vertex is given a spin (up or down depending on the sign). It means, that each state has a number of frustrated edges. The overall energy of the system increases with the number of frustrated edges. Thus, the state with the least energy, the ground state, is the one that minimizes the frustration index.

3.2. Polarity Opinion Graph

In the particular case of a SM network, V represents the set of users and E the interactions and/or relationships between users.

Depending on the particular SM network, the edges can represent direct relationship between users (u_i follows u_j or u_i is a friend of u_j) or interaction-based relationship between users. Given a SM interaction t_k so that user u_i has authored it ($author(u_i, t_k)$ *is True*), there is an additional SM interaction t_l authored by a different user u_j, so that t_l represents a reaction of u_j to the interaction t_k. This reaction is typically a *like*, a *share*, a *reply*, a *comment*, etc depending on the particular SM platform.

A SM Interaction represents the atomic piece of content generated by the user u_i during the time $\triangle\, t$ in a SM platform (e.g.: a tweet, a Facebook post, a LinkedIn status update, etc). Thus, $Author(u_i, it_k, \triangle\, t)$ is a function that retrieves *True* if u_i created the interaction it_k in the time period $\triangle\, t$, and *False* otherwise. The time interval $\triangle\, t$ might be measured in weeks, days or hours, depending on the use case and consists of two extremes: $t_{startdate}$ and end date $t_{enddate}$.

We define the set of user interactions It for a given user u_i within a time interval $\triangle\, t$, as: $It(u_i, \triangle\, t) \equiv \{it\}, \forall it_k Author(u_i, it_k, \triangle\, t)$

To determine if a particular interaction it_k belongs to a topic T, we compute the *semantic relatedness* between the content of the interaction and the semantic representation of the topic T and agree on a threshold. If is above a particular threshold Θ

$$InTopic(it_k, T) = \begin{cases} True & \text{if } semanticrelatedness(it_k, T) > \Theta \\ False & otherwise \end{cases} \tag{3}$$

We define a **signed polarity interaction** as any SM engagement of a particular user u_i with any particular content item created by any other user u_j, where $u_i \neq u_j$. Signed polarity interactions can be positive or negative to different degrees, representing to which extent user u_j supports or oppose the content of the interaction. Depending on the possibilities of engagement offered by the SM platform, the sign of the iteration might be more or less immediate to infer beyond the content of the interaction (e.g.:

In Twitter, interactions can be *retweets* -which usually indicates strong agreement-) We define $polarEngagement(u_k, u_j, it_i)$ as follows:

$$polarEngagement(u_k, u_j, it_i) = \begin{cases} -1 & \text{if } Author(u_i, it_i)\ \exists it_j, Author(u_j, it_j),\ contradicts(i_i, i_j) \\ 1 & \text{if } Author(u_i, it_i)\ \exists it_j, Author(u_j, it_j)\ supports(i_i, i_j) \\ 0 & otherwise \end{cases} \tag{4}$$

Both *contradicts* and *supports* are logical conditions established as a partition space of the semantic function $agreement(i_i, i_j) : E \rightarrow [-1, +1]$, which measures the semantic level of agreement or disagreement between the interactions it_i and it_j. Even if a Likert-like partitioning could apply, we rather consider the partitions at the extremes to enable the creation of the polarity graph, as we will discuss below. Thus, the two partitions are defined as follows:

$$contradicts(i_i, i_j) = \begin{cases} True & \text{if } agreement(i_i, i_j) < \theta \\ False & \text{otherwise} \end{cases} \quad with\ \theta < 0 \tag{5}$$

and

$$supports(i_i, i_j) = \begin{cases} True & \text{if } agreement(i_i, i_j) > \vartheta \\ False\&\text{otherwise} \end{cases} \quad with\ \vartheta < 0 \tag{6}$$

Both are governed by two thresholds, one negative θ, and one positive ϑ to define the size of the partition and to prevent overlaps.

The agreement function can be implemented using methods from the argument mining field [38,39]. It can also be seen as a Natural Language Inference problem and resolved by state-of-the-art methods, such as [40]. *Agreement detection* is a research field in its own right which would go beyond the scope of this paper.

Polarity disambiguation is the process of consolidating several interactions $It \equiv it_1, ..., it_n, \forall x, x \in 1, .., n, polarEngagement(u_i, u_j, it_x) \neq 0\, OR\, polarEngagement(u_j, u_i, it_x) \neq 0$ between users u_i and u_j into a single signed value. Polarity disambiguation refers to the process of deciding whether in the interactions between u_i and u_j are predominantly in agreement or in disagreement.

$$polarityDisambiguation(u_i, u_j, It) \rightarrow G : (V = \{u_i, u_j\}, E = (\overline{u_i, u_j}), \sigma = \sigma_{dis}). \tag{7}$$

The **Opinion Polarity Graph** is a signed non-complete graph to reflect the nature of the interactions between a set of users given a particular topic over a period of time. The algorithm Alg. 1 explains how the *Opinion Polarity Graph* is built.

Fig. 2 shows an example about our proposed method to derive the *Polarity Opinion Graph*. Light-gray shaded interactions belong to the given Topic T. $It_{12}, It_{k3}, It_{24}$ and It_{46} are original posts. All others are reactions to the original posts. It_{23} represent the disagreement of u_2 and u_2 with regards to it_{12}, which translates into a negative edge between u_1 and u_2 in the resulting *Polarity Opinion Graph*. It_{k3} has been created by u_5 before the analysis period p, but as the user u_4 expressed agreement in the interaction it_{42}, which takes place within the period p, we consider the user u_5 as part of the resulting graph with a positive connection to u_4. u_3 has responded once positively to one post of u_1 (see it_{33} to it_{13}) and once with disagreement (see it_{35} to it_{1l}). In this case, we need to apply a polarity disambiguation procedure to determine the sign between u_3 and u_1 Posts with no reaction, such as It_{3s}, does not have any effect in the resulting *Polarity Opinion Graph*

```
Data: All interactions in the defined period related to the topic
      $It \equiv \{it_i\}, InTopic(it_i, T)$
Data: All users who authored these interactions
      $U \equiv \{u_k\}, \forall it_i \in It, Author(u_k, it_i)$
Result: Opinion Polarity Graph G
$G@V = \varnothing$ ;
$G@E = \varnothing$ ;
for $u_k \in U$ do
    for $u_j \in U$ do
        if $\exists it_1, polarEngagement(u_k, u_j, it_i) \neq 0$ then
            $sign \longleftarrow polarityDisambiguation_i^n(polarEngagement(u_k, u_j, it_i),$
                $\forall it_k, polarEngagement(u_k, u_j, it_k) \neq 0$
            $G@V \longleftarrow G@V \cap u_k$ (adding the user $u_k$ as node) ;
            $G@V \longleftarrow G@V \cap u_j$ (adding the user $u_j$ as node) ;
            $G@E \longleftarrow G@E \cap edge(u_k, u_j)$ (adding an edge between users) ;
            $\sigma(u_k, u_j) \longleftarrow sign$ (defining the adding an edge between users) ;
        end
    end
end
```

Algorithm 1: Algorithm to build the Polarity Opinion Graph

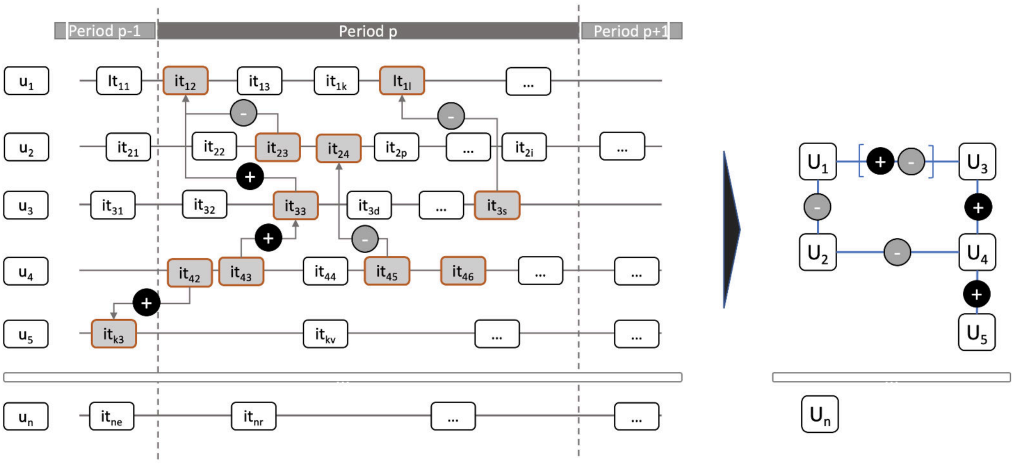


Figure 2. Polarity Opinion Graph building example

In Fig. 3 we show a real example on how the interactions between three users A, B and C translate into a particular set of vertexes, edges and polarities in a larger *Opinion Polarity Graph*.

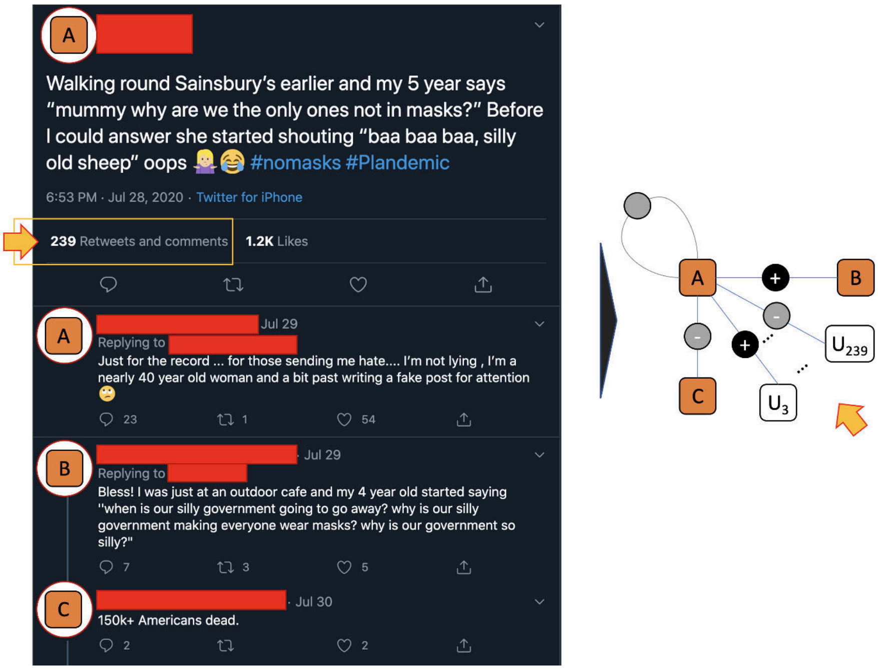
3.3. Structural Imbalance analysis

Once we have the Polarity Opinion Graph defined, we need a method to extract the set of edges that violate the structure balance.

We define $SB_{violations} = (V', E', \sigma)$ as the collection of signed edges between vertexes that represent violations of the balanced state of our Polarity Opinion Graph G

The SB problem can be mapped onto the Quantum Annealer objective functions in order to solve via quantum annealing. The SB problem maps to an Ising formulation, as discussed in the subsection 2.3. The coefficient matrices for SB is provided below:

$$J_{ij}^{SBP} = \begin{cases} sign(it_i, it_j) & \text{if } (it_i, it_j) \in E \; and \; i < j \\ 0 & otherwise \end{cases} \tag{8}$$



A [redacted]

Walking round Sainsbury's earlier and my 5 year says "mummy why are we the only ones not in masks?" Before I could answer she started shouting "baa baa baa, silly old sheep" oops 🤷‍♀️😂 #nomasks #Plandemic

6:53 PM · Jul 28, 2020 · Twitter for iPhone

239 Retweets and comments 1.2K Likes

A [redacted] Jul 29
Replying to [redacted]
Just for the record ... for those sending me hate.... I'm not lying , I'm a nearly 40 year old woman and a bit past writing a fake post for attention 🙄
23 1 54

B [redacted] · Jul 29
Replying to [redacted]
Bless! I was just at an outdoor cafe and my 4 year old started saying "when is our silly government going to go away? why is our silly government making everyone wear masks? why is our government so silly?"
7 3 5

C [redacted] · Jul 30
150k+ Americans dead.
2 2

Figure 3. Example of a social media interaction translating into the Polarity Opinion Graph

3.4. Moderating opportunities

Once we have identified the violating edges and correspondingly the users participating in these interactions $SB_{violations} = (V', E', \sigma)$, we need to define a set of actions to balance the graph.

In graph theory, options such as switching or deleting edges ([37], which are commonly offered to restore the structural balance, are not practicable in the SM context. Deleting a particular interaction cannot be technically done by any other user than the author (and the SM administrators). In addition, it is non-realistic to ask users to remove their posts.

Rather, we need to find ways to minimize the structural imbalance in a *non-intrusive* way (using engagement approaches any user could implement without needing additional privileges -super-user, admin, designated moderator, etc-). Intuitively, the idea is to trigger the creation of further interactions with the opposite polarity between any pair of users in $SB_{violations}@V'$. In Fig. 4 we show one approach , where a *moderating interaction* by a third user K which mentions A and B might potentially trigger a new interaction between A and B with opposite sign. If we apply *polarity Disambiguation*, we can consider the balance violation given by the polarity of the edge between A and B eventually solved.

4. System Overview

In this section, we describe the design and implementation of the system to enable the discovery and the moderation of the social debate in a SM network to optimize the effectiveness of each moderation attempt, as described in the previous section.

Fig. 5 shows the main building blocks and the typical flow of execution:

1. The *SM harvester* in combination with the *Topic Modeller* pull all relevant content related to a particular topic given a set of seed hashtags as input, following the semantic expansion approach described in [41] to discover and fetch all relevant content related to the topic.
2. For all the topic-related posts, all related interactions (responses, re-shares, etc) are also gathered, alongside all information and properties related to the users authoring either original posts or responds / re-shares.

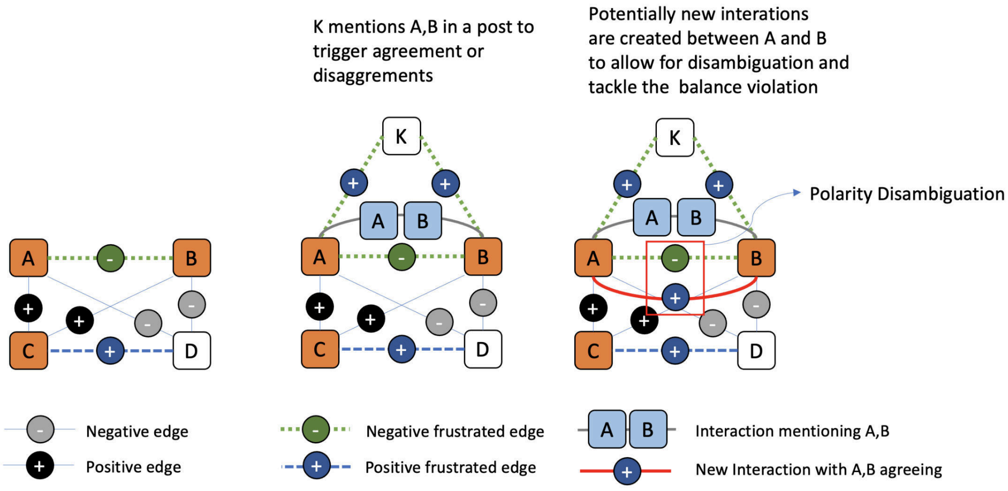

K mentions A,B in a post to trigger agreement or disaggrements
Potentially new interations are created between A and B to allow for disambiguation and tackle the balance violation
Polarity Disambiguation
K
A
B
C
D
Negative edge
Positive edge
Negative frustrated edge
Positive frustrated edge
Interaction mentioning A,B
New Interaction with A,B agreeing

Figure 4. Moderation opportunities

3. Applying the algorithm 1, described in the previous section, to all gathered content and users, the Polarity Opinion Graph is built.
4. The Polarity Opinion Graph (or any of logical partition, for example, by day, by hashtag, etc, depending on the pursued moderation in the analysis) is then submitted to the Structural Balance Solver in the Quantum Annealing system, which returns the set of edges that violates the structural balances of the graph (if any)
5. The non-intrusive moderation takes place based on the set of interactions cause imbalance. A human moderator needs to decide the best content to engage with the users involved in the violating interactions.
6. Once we register any new interaction between any of the users involved in a balance violation, the polarity engagement of this interaction is computed applying equation 3.2 followed by the polarity disambiguation defined in the equation 3.2 between the violating interaction and the new one, which might lead to a new polarity between the involved users. If we then recompute the *Polarity Opinion Graph* and rerun the *Structural Balance Solver*, a new frustration index value will reflect the effect of the moderations.

4.1. Modules description

Once we have described the flow, we will describe the different modules of the system and comment on their potential implementations:

4.1.1. Social Media Harvester

The purpose of this module is to interact with the SM network, retrieving the required content about interactions, users, trending topics, etc.

It can perform searches for a particular term or hashtag and pull all interactions tagged with this term. In addition, it can fetch all information related to a particular user or set of users, including followers, friends, interactions, etc. Additional filters, such as language, location, etc to make the search universe more targeted.

Typically, this module connects with REST or streaming APIs from the SM platforms. It is important to understand if the SM platform retrieves the whole result set for a query or just a sample and in the latter case, which are the criteria to select the records within the sample. It might affect the signed graph.

Strategies often recommend using some metrics, such as user centrality (based on the number of connections, followers/friends, etc), exposure, activity index, [42], etc. to

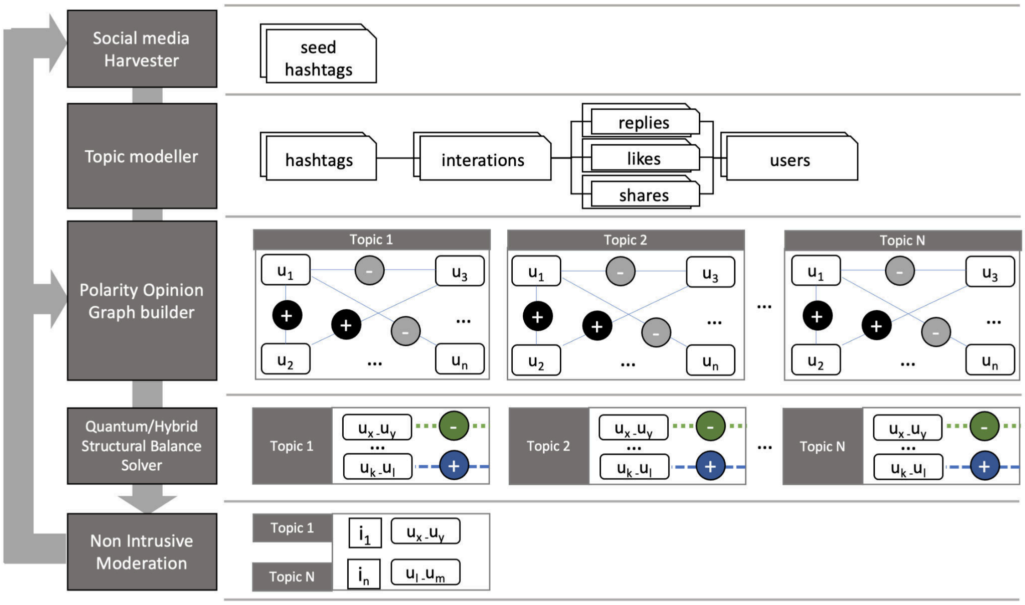


Figure 5. Overview System

focus on the users with the highest potential to shape the SM dialogue vs. any random sampling criteria implemented in the SM platform API.

4.1.2. Topic modeller

This module steers the SM harvester to build up the set of topics under analysis. A proper definition of a topic is key to uncover moderation opportunities within the same context and thus, increasing the chance of users to engage with moderating interactions.

This module interacts with the harvester to implement the semantic expansion algorithm based on the discovery of new hashtags given a set of seed hashtags, optimizing co-occurrence and frequency, as explained in [41]

The *Topic modeller* deals with the implementation of the semantic relatedness to determine if a particular interaction can be labelled as belonging to the topic or not, as defined in equation 3.2. There are several approaches to implement this function, as explained in [43,44]

4.1.3. Polarity Opinion Graph builder

The information gathered by the previous 2 modules is applied the algorithm 1 to create the Polarity Opinion Graph.

The process builds upon 2 natural language processing functions: a) *polarity Engagement* (see equation 3.2) to defined if an interaction supports or contradicts another interaction and derive the sign of the edge between the involved users and b) polarity disambiguation, which follows the algorithm Alg. 1

As expressed in the subsection 3.2, it is important to implement a polarity disambiguation schema that best reflects the intent of the moderation. There are different possibilities to implement this concept: a) considering the polarity from the most recent interaction as leading polarity, b) choosing the polarity of the interaction with the maximum absolute value for the level of agreement, c) combining recency and level of agreement, etc.

4.1.4. Quantum/Hybrid structural balance solver

As explained in the background section, quantum annealing is a feasible technology to solve the NP-hard task of finding the edges violating the structural balance in a signed graph.

In [18] the authors provide the proof for the superior performance reached in the quantum mechanical processes implementation. The D-wave quantum annealing hardware outperformed all similar approaches based on software simulation to solve these tasks.

The D-Wave implementation minimizes the following Ising objective function.

$$O(h, J, s) = \sum_{i} h_i s_i + \sum_{i<j} J_{ij} s_i s_j \tag{9}$$

where spin variables $s_i \in -1, +1$ are subject to local fields and pairwise interactions with coupling strengths J_{ij}.

Unlike classical computers, Quantum hardware uses qubits instead of bits to hold information. Qubits are governed by the quantum mechanisms laws, which allows them for being in *superposition* (both are -1 and +1 at the same time) until they collapse to either -1 or +1 triggered by an outside event.

An annealer produces as output a low energy ground state s, which consists of an Ising spin for each qubit where $s_i \in -1, +1$. D-Wave and other quantum annealers are built upon this basis and support the solving of NP-hard problems (including optimization, sampling, machine learning, etc), as demostrated in [18].

The D-Wave Quantum Annealer uses the so called Chimera graph to represent both the connectivity between binary variables s_i. Thus, in the Chimera graph $G = (V, E)$, nodes in V as qubits s_i and s_j represent variables of the problem to solve with programmable weights h_i, and edges in E, couplers for the nodes with programmable strength J_{ij}.

4.1.5. Non intrusive moderation

At present, this part is for a human editor to take over, as explained in section 3.3. It requires trained moderation skills and experience to trigger the reaction of any of the users involved in an interaction discovered by the previous module.

As part of future research, we foresee the use of recommender techniques to assist the human moderator, but we don't foresee a fully automated approach. In addition, moderation might require deep subject matter expertise in the topic, which renders any automation attempt unrealistic.

5. Experimental results

In order to discuss the performance of our proposed model, we have implemented the system for the SM network Twitter.

We select the well-known and presently very relevant topic of *"COVID-19"*, because of following reasons: a) the high volume of interactions, b) the richness of encountered opinions, c) the degree of overlapping with other heated political debates (e.g.: Russian role on the US elections, Trump's controversy, anti-vaccination movements, etc.), d) the global nature of the debate.

We focused on 10 days (21^{st} to the 31^{st} of July in 2020) and gathered 121023 interactions (tweets, replies, re-tweets) in English related to the *COVID-19*.

To harvest topic-specific only content, we followed the semantic expansion approach we shared in the subsection 4.1, which starts with a set of user-defined hashtags acting as seeds and while fetches integrations related to these seeds, discovers further seeds.

The seed hashtags we selected are: *#coronavirus, #antimask, #covidiots, #nomask, #covidhoax, #scamdemic, #plandemic*.

In table 1, we show the number of interactions per hashtag as well as the top hashtags sorted by the number of occurrences. While we observe solid consistency with regards to the selected topic, there are some hashtags that could be grouped as a subtheme within the topic COVID-19 or as a subtopic, such as the ones related to Russia, etc. Our approach would allow for running the same analysis at subtopic level.

Table 1: Topic related hashtags and occurrences

Hashtag	Occur.	Hashtag	Occur.	Hashtag	Occur.
NoMasks	5600	MoscowMitch	2198	NoMaskSelfie	768
TrumpVirus	5371	Russia	1844	CovidHoax	721
NoMask	4959	KBF	1823	Covidiots	700
coronavirus	4036	NobodyLikesTrump	1587	FauciTheFraud	645
FoxBusiness	3895	Plandemic	1503	londonprotest	624
FoxNews	3889	covidiots	1027	RampageRantz	623
nyc	3887	covid	938	science	584
COVIDIOTS	3373	USA	937	Norwich	559
Tulsa	2753	karenfrombunnings	824	covid19hoax	558
COVID19	2660	bunnings	819	COVID__19	547

To build up the Polarity Opinion Graph, we follow the algorithm described in Alg. 1. We use the engagement mode offered by Twitter to interact with the post of a particular user: *Reply* and *Retweet* (we refrained from using *Like*, as it is not possible to act or disambiguate and it doesn't represent a real dialogue). We have interpreted "retweeting" as a supporting link. In the case of replying, we have extracted the sentiment and only if the sentiment is highly positive, we interpret it as "supporting"; in any other case we see it as "contradicting". We computed using the Syuzhet dictionary, but could have been any conventional sentiment analysis approach [45].

Fig. 6 shows a circular representation of the polarity opinion graphs for the top hashtags over the 10 days of analysis. Users are represented by orange circles whose size is proportional to the number of followers. Positive edges are represented in dark blue while negative ones in light blue. When the number of links is very high, the links of a color might prevent the links from the other color to be visible, like in the case of *#trumpvirus*.

After obtaining the polarity opinion graph, we can map the problem to the Quantum Annealing DW-2000Q_6, from the manufacturer D-Wave. This QPU is fabricated with 2048 qubits and 6016 couplers in a Chimera topology, as described in the standard manual [1]. The mapping and the solving are performed using the SDK and the libraries provided for python clients by D-Wave itself, abstracting the complexity of low-level embedding into the QPU.

Depending on the needs for moderation, we can filter by hashtag to narrow the scope and we can also filter by period. In Fig.7, we show the daily structural balance violations (in orange) in 6 different consecutive days for the interactions related to the hashtag *#nomask*. Each and every orange edge represents an opportunity for non-intrusive moderation.

In Table 2, we see the results for graph cuts based on the occurrence of a particular hashtag. We observe that this constraint massively limits the number of edges. If we don't filter by hashtag and rather run the analysis with the complete polarity opinion graph, containing 22899 nodes and 24851 edges, a total of 647 structural balance violations are detected.

Out of the 647 violations, we moderated 200 (31%), obtaining 46 reactions (23%). With polarity disambiguation applied to give the full weight to the polarity of the latest

[1] https://docs.dwavesys.com/docs/latest/c_gs_4.html

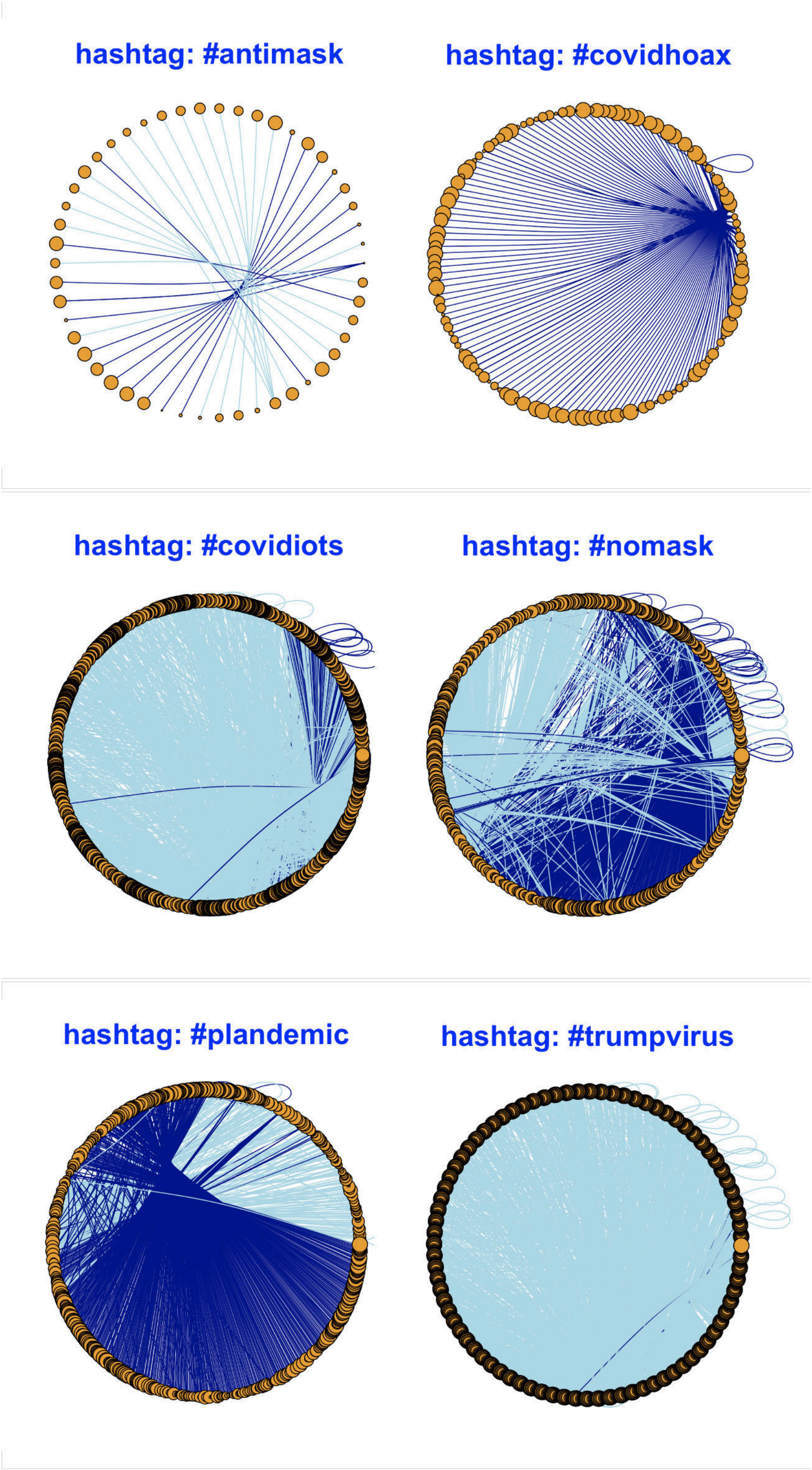


Figure 6. Polarity Opinion Graphs for top hashtags in the COVID-19 topic between 11th and 21st of July 2

interaction, we managed to correct 12 violations (27%). A funnel view can be seen in Fig. 8

The moderation has been performed out of the same account and using an interrogative style (one example in Fig. 9). There are several factors that might impact the success rate of the total moderations, such as quality of the copy, use of more or less aggressive moderation styles, use of rich content (videos, pictures) or hyperlinks, etc.

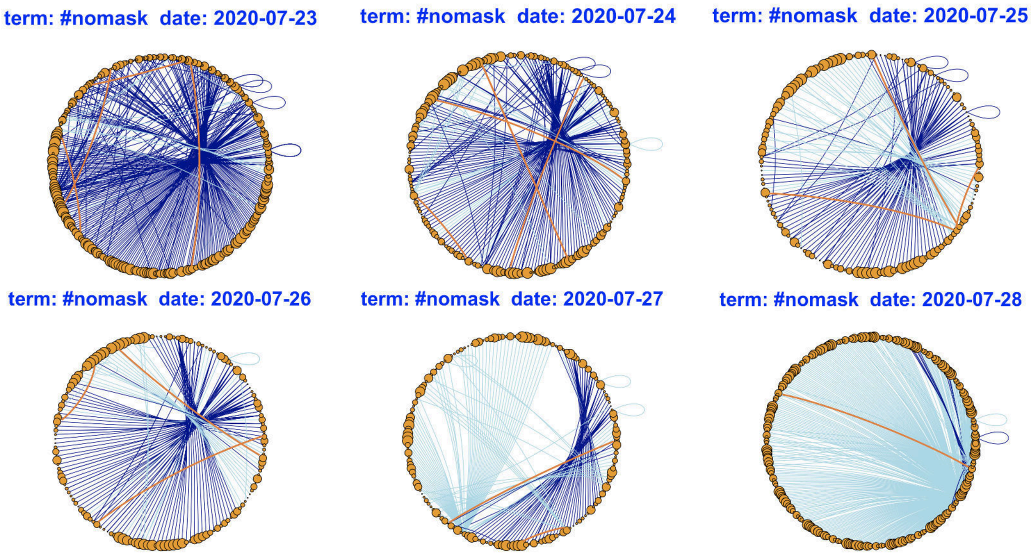

term: #nomask date: 2020-07-23
term: #nomask date: 2020-07-24
term: #nomask date: 2020-07-25
term: #nomask date: 2020-07-26
term: #nomask date: 2020-07-27
term: #nomask date: 2020-07-28

Figure 7. Polarity Opinion Graphs for top hashtags in the COVID-19 topic between 11^{th} and 21^{st} of July 2020 with balance violations in orange

Table 2: Experiment summary

Hashtag	Nodes	Edges	Violations	Hashtag	Nodes	Edges	Violations
#nomask	3485	425	5	#covidhoax	618	140	2
#trumpvirus	4321	1579	1	#plandemic	2153	548	5
#scamdemic	1297	159	0	#covidiots	4370	678	0

Given the variety of techniques and the complexity of evaluation, we suggest exploring that in future research.

We have focused for this paper on moderations seeking to re-establish the structural balance of the polarity opinion graph, but depending on the purpose of the moderating party, our method can also be targeted to just moderating the disagreements, or the agreements or to systematically influence the overall SM dialogue on a particular topic.

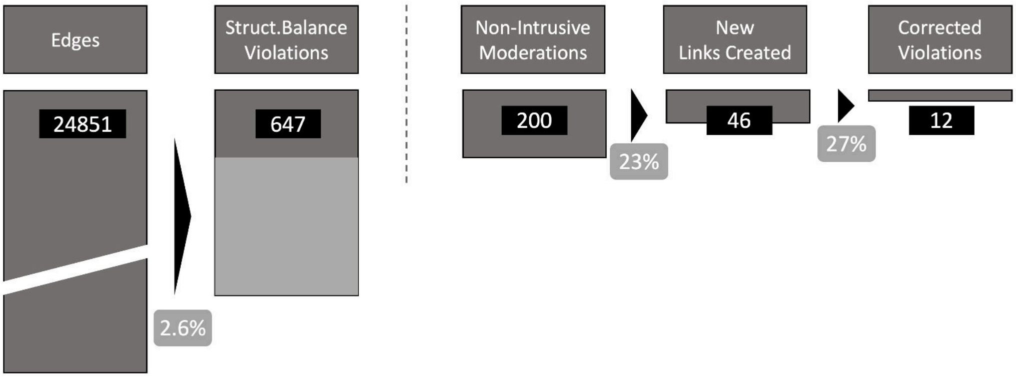


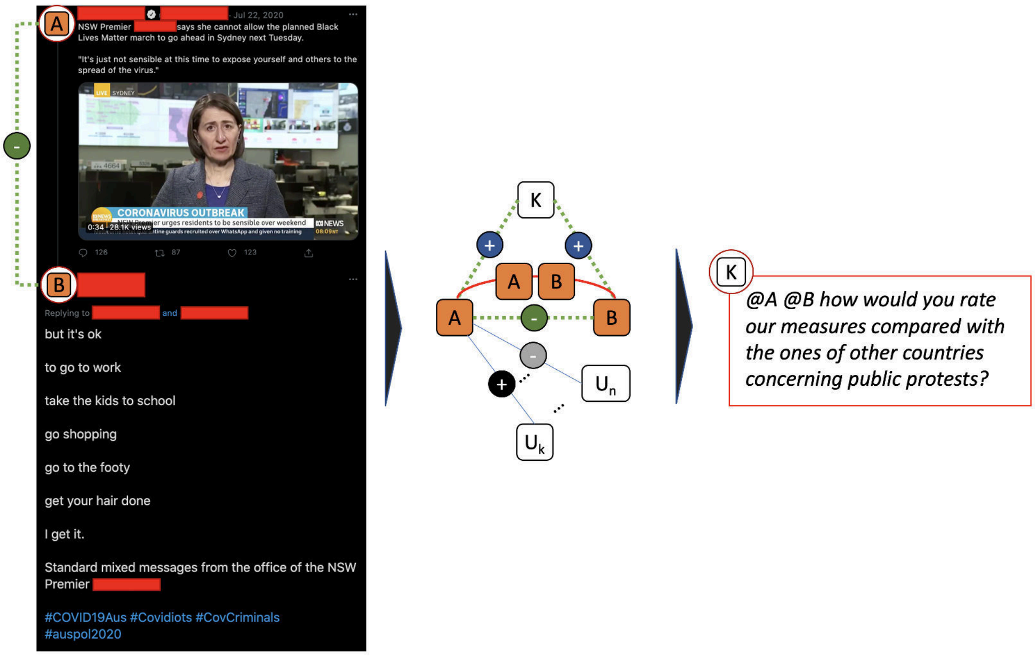
Figure 8. Funnel showing the effect of the non-intrusive moderations

The full experiment shows how our method works applied to the "COVID-19" topic, with the setup and the results summarized in Table 3. The use of Quantum annealing guarantees the feasibility of identifying the structural balance violations for large polarity opinion graphs. We observe also that non-intrusive moderations alleviate the balance problem and help therefore balancing the SM dialogue for the topic. At the same time, the experimental results show a whole set of possibilities to adapt our

Table 3: Experiment summary

Period	21^{st}- 31^{st} July 2020	Location	Worldwide
#Tweets	121023	#Re-tweets	73632
#Replies	47401	LanguageFilter	EN
#Edges Opinion Graph	24851	#Nodes Opinion Graph	22899
#Violations	647	#Moderations	200
#New Links	46	#Corrected Violations	12

method to subtasks (such as daily moderation, hashtag-based moderation, sub-topic moderation, etc). Also, worth-while mentioning are the implementation decisions we have taken for this experiment, which have an impact on the overall performance; for instance, the method for the semantic relatedness to asses which interaction belongs to the topic or the method to define if an interaction supports or contradicts another interaction, or the moderation style, as commented above.



[redacted] [redacted] · Jul 22, 2020

NSW Premier [redacted] says she cannot allow the planned Black Lives Matter march to go ahead in Sydney next Tuesday.

"It's just not sensible at this time to expose yourself and others to the spread of the virus."

126 87 123

[redacted]

Replying to [redacted] and [redacted]

but it's ok

to go to work

take the kids to school

go shopping

go to the footy

get your hair done

I get it.

Standard mixed messages from the office of the NSW Premier [redacted]

#COVID19Aus #Covidiots #CovCriminals #auspol2020

@A @B how would you rate our measures compared with the ones of other countries concerning public protests?

Figure 9. Example of a moderation to trigger a new positive link between A and B from the 21^{st} of July 2020

6. Conclusion

In this paper, we have introduced a novel methodology to define a SM moderation strategy around a specific topic based on the principle of structural balance and using non-intrusive techniques. Our methodology relies on mining all relevant interactions related to the defined topic, alongside the reactions in the SM related to these interactions and the users involved. We achieved that defining a topic membership function based on semantic relatedness and we grow the topic by applying a well-defined semantic expansion technique.

We introduced the concept of *polarity engagement* based on 2 NLP functions to determine if the response to a SM interaction supports or contradicts the original interaction, which translates into a positive or negative polarity between the SM users involved. In addition, our work defines the concept of *polarity disambiguation* to determine the leading polarity between two users, when there is more than 1 interaction between these 2 users over a period of time. Based on these concepts, we provide an algorithm to build up the so-called *polarity opinion graph*, which represented the disambiguated polarity between all users sharing interactions for the topic under analysis.

The moderation opportunities are defined as the violations of the structural balance in the *polarity opinion graph*, which we compute mapping the problem to a Quantum Annealer QPU to guarantee the feasibility given the NP-hard nature of the problem with classic computers.

After identifying the moderation opportunities, we then suggest a non-intrusive approach to trigger the creation of new links between the involved users to neutralize the polarity of the existing ones and therefore contribute to the overall balance reduction.

In addition to formalizing our approach, we provide the overview of a system to implement it, as well as the description of the components, their technical foundation and how they are orchestrated to achieve the information flow required in our approach.

To make a proper assessment of our methodology, we implemented the system integrating with Twitter and focused on the COVID-19 topic, harvesting SM interactions for 10 days. In the experiment we gathered more than 121K tweets, 73K re-tweets and 47K replies resulting in a polarity opinion graph with ca. 23K nodes and 24.5K signed edges. We used a DW-2000Q_6 Quantum Annealer from D-Wave to extract the violations of the structural balance (647 in total) and focused on 200 for applying non-intrusive moderation. It triggered the creation of 46 new links and solved 12 violations (6%), but also showed further moderation possibilities that are feasible with our methodology.

In addition to the originality of the methodology just described, this work introduces several novel concepts, such as *polarity engagement*, *polarity disambiguation* in the context of SM and provides a new algorithm to build the *Opinion polarity graph* from a set of interactions and users in a SM network. To add to the definition of non-intrusive moderation based on the structural balance violations encountered by the quantum annealer.

Further research work could focus on the benchmarking of different non-intrusive moderation strategies, as it can positively affect the overall performance. Our approach could also profit from relationship mining techniques between the users without the need for an explicit response by, for instance, leveraging the concept of exposure. An additional idea connected to this research direction would consider information diffusion schemas for both initial posts and responds, resulting in dynamic polarity opinion graphs. Last but not least, we see the potential for moderation techniques pursuing the maximum structural imbalance and also the detection of these manipulative attempts.

References

1. Jain, L.; Katarya, R. Discover opinion leader in online social network using firefly algorithm. *Expert Systems with Applications* **2019**, *122*, 1–15.
2. Stieglitz, S.; Dang-Xuan, L. Social media and political communication: a social media analytics framework. *Social network analysis and mining* **2013**, *3*, 1277–1291.
3. Drummond, C.; O'Toole, T.; McGrath, H. Digital engagement strategies and tactics in social media marketing. *European Journal of Marketing* **2020**.
4. Briggs, S. The Freedom of Tweets: The Intersection of Government Use of Social Media and Public Forum Doctrine. *Colum. JL & Soc. Probs.* **2018**, *52*, 1.
5. Bradshaw, S.; Howard, P. Troops, trolls and troublemakers: A global inventory of organized social media manipulation **2017**.
6. Wu, L.; Morstatter, F.; Carley, K.M.; Liu, H. Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter* **2019**, *21*, 80–90.
7. Berghel, H. Malice domestic: The Cambridge analytica dystopia. *Computer* **2018**, pp. 84–89.
8. Allcott, H.; Gentzkow, M. Social media and fake news in the 2016 election. *Journal of economic perspectives* **2017**, *31*, 211–36.
9. Van der Linden, S.; Leiserowitz, A.; Rosenthal, S.; Maibach, E. Inoculating the public against misinformation about climate change. *Global Challenges* **2017**, *1*, 1600008.
10. Tambuscio, M.; Oliveira, D.F.; Ciampaglia, G.L.; Ruffo, G. Network segregation in a model of misinformation and fact-checking. *Journal of Computational Social Science* **2018**, *1*, 261–275.
11. Bail, C.A.; Argyle, L.P.; Brown, T.W.; Bumpus, J.P.; Chen, H.; Hunzaker, M.F.; Lee, J.; Mann, M.; Merhout, F.; Volfovsky, A. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9216–9221.
12. Jiang, J.A. Identifying and Addressing Design and Policy Challenges in Online Content Moderation. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–7.

13. Roberts, S.T. *Behind the screen: Content moderation in the shadows of social media*; Yale University Press, 2019.
14. Askarisichani, O.; Lane, J.N.; Bullo, F.; Friedkin, N.E.; Singh, A.K.; Uzzi, B. Structural balance emerges and explains performance in risky decision-making. *Nature communications* **2019**, *10*, 1–10.
15. Summers, T.H.; Shames, I. Active influence in dynamical models of structural balance in social networks. *EPL (Europhysics Letters)* **2013**, *103*, 18001.
16. Altafini, C. Dynamics of opinion forming in structurally balanced social networks. *PloS one* **2012**, *7*, e38135.
17. Baker, S.A.; Wade, M.; Walsh, M.J. <? covid19?> The challenges of responding to misinformation during a pandemic: content moderation and the limitations of the concept of harm. *Media International Australia* **2020**, *177*, 103–107.
18. Bian, Z.; Chudak, F.; Macready, W.G.; Rose, G. The Ising model: teaching an old problem new tricks. *D-wave systems* **2010**, *2*.
19. Loader, B.D.; Mercea, D. Networking democracy? Social media innovations and participatory politics. *Information, Communication & Society* **2011**, *14*, 757–769.
20. Yang, X.; Chen, B.C.; Maity, M.; Ferrara, E. Social politics: Agenda setting and political communication on social media. International Conference on Social Informatics. Springer, 2016, pp. 330–344.
21. Russell Neuman, W.; Guggenheim, L.; Mo Jang, S.; Bae, S.Y. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication* **2014**, *64*, 193–214.
22. Khan, J.Y.; Khondaker, M.; Islam, T.; Iqbal, A.; Afroz, S. A benchmark study on machine learning methods for fake news detection. *arXiv preprint arXiv:1905.04749* **2019**.
23. Gillespie, T. Content moderation, AI, and the question of scale. *Big Data & Society* **2020**, *7*, 2053951720943234.
24. Link, D.; Hellingrath, B.; Ling, J. A Human-is-the-Loop Approach for Semi-Automated Content Moderation. ISCRAM, 2016.
25. Yadin, D.; Yahav, I.; Zalmanson, L. The Guide to Content Moderation: Introducing Crowds to Mitigate the Challenges of the Human Moderator **2020**.
26. Myers West, S. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* **2018**, *20*, 4366–4383.
27. Leskovec, J.; Huttenlocher, D.; Kleinberg, J. Signed networks in social media. Proceedings of the SIGCHI conference on human factors in computing systems, 2010, pp. 1361–1370.
28. Rawlings, C.M.; Friedkin, N.E. The structural balance theory of sentiment networks: Elaboration and test. *American Journal of Sociology* **2017**, *123*, 510–548.
29. Antal, T.; Krapivsky, P.L.; Redner, S. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena* **2006**, *224*, 130–136.
30. Kunegis, J. Applications of structural balance in signed social networks. *arXiv preprint arXiv:1402.6865* **2014**.
31. Heider, F. Attitudes and cognitive organization. *The Journal of psychology* **1946**, *21*, 107–112.
32. Aref, S.; Wilson, M.C. Balance and frustration in signed networks. *Journal of Complex Networks* **2019**, *7*, 163–189.
33. Aref, S.; Mason, A.J.; Wilson, M.C. A modeling and computational study of the frustration index in signed networks. *Networks* **2020**, *75*, 95–110.
34. Friedli, S.; Velenik, Y. *Statistical mechanics of lattice systems: a concrete mathematical introduction*; Cambridge University Press, 2017.
35. Facchetti, G.; Iacono, G.; Altafini, C. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences* **2011**, *108*, 20953–20958.
36. Wang, S.; Gong, M.; Du, H.; Ma, L.; Miao, Q.; Du, W. Optimizing dynamical changes of structural balance in signed network based on memetic algorithm. *Social Networks* **2016**, *44*, 64–73.
37. Harary, F.; others. On the notion of balance of a signed graph. *Michigan Mathematical Journal* **1953**, *2*, 143–146.
38. Dusmanu, M.; Cabrio, E.; Villata, S. Argument mining on Twitter: Arguments, facts and sources. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2317–2322.
39. Cocarascu, O.; Toni, F. Combining deep learning and argumentative reasoning for the analysis of social media textual content using small data sets. *Computational Linguistics* **2018**, *44*, 833–858.
40. Sun, Z.; Fan, C.; Han, Q.; Sun, X.; Meng, Y.; Wu, F.; Li, J. Self-Explaining Structures Improve NLP Models. *arXiv preprint arXiv:2012.01786* **2020**.
41. Bernabé-Moreno, J.; Tejeda-Lorente, Á.; Herce-Zelaya, J.; Porcel, C.; Herrera-Viedma, E. An automatic skills standardization method based on subject expert knowledge extraction and semantic matching. *Procedia Computer Science* **2019**, *162*, 857–864.
42. Bernabé-Moreno, J.; Tejeda-Lorente, A.; Porcel, C.; Fujita, H.; Herrera-Viedma, E. CARESOME: A system to enrich marketing customers acquisition and retention campaigns using social media information. *Knowledge-Based Systems* **2015**, *80*, 163–179.
43. Taieb, M.A.H.; Zesch, T.; Aouicha, M.B. A survey of semantic relatedness evaluation datasets and procedures. *Artificial Intelligence Review* **2020**, *53*, 4407–4448.
44. Gupta, A.; Kumar, A.; Gautam, J. A survey on semantic similarity measures. *International Journal for Innovative Research in Science & Technology* **2017**, *3*, 243–247.
45. Yoon, S.; Parsons, F.; Sundquist, K.; Julian, J.; Schwartz, J.E.; Burg, M.M.; Davidson, K.W.; Diaz, K.M. Comparison of different algorithms for sentiment analysis: Psychological stress notes. *Studies in health technology and informatics* **2017**, *245*, 1292.