Department of Mathematics and Statistics

Weighted -rate dominating sets in social networks

Danica Vukadinovic Greetham¹, Anush Poghosyar¹, and Nathaniel Charlton^{1,2}

¹ Centre for the Mathematics of Human Behaviour Department of Mathematics and Statistics University of Reading, UK In health-related behaviour change context, for an intervention to work at the individual level, it is often of the utmost importance that a support network exist (see e.g. [8]). In this way an individual is surrounded with social support. Also, a support network needs to have a major in uence on the individual, as possible negative in uences also come from her/his social network (for example in interventions aimed at addictive behaviours).

For these reasons, one often needs to nd a set of nodes/individuals such that all other or indeed all individuals are connected to that set. In graph theory such a set is called a dominating set and a problem of nding a dominating set of minimal cardinality is NP complete [7]. The notion was generalised introducing k-domination where each node needs to have at least neighbours in the dominating set, and domination where 0 < 1, where each node not in the dominating set needs at least 100 percentage of neighbours in the dominating set [12]), and -rate domination [6] where each node (including ones in the dominating set) needs to have at least 100 percentage of neighbours in the dominating set. Again, nding minimum cardinalities of and -rate dominating sets is NP-complete.

Here, we introduce -rate dominating sets problems on weighted networks. Why weighted networks? It might be that the \best" candidates (from structural perspective) for dominating sets are not feasible for di erent reasons: they cannot be a part of intervention because they do not have desired attributes, or they do not have time to invest into intervention. We want to overcome this assigning a cost to be part of intervention to each node. Thus, our goal is to nd a most cost e ective set that we can control or dominate network from. Note that here we do not model negative in uences that come from a social network, but just require at least 100 percents of neighbours to be in the support network.

In the next section we give preliminaries and formally de ne the problem. In Section 2 an overview of the previous work is given. In Section 3 theoretical upper bound on weighted -rate dominating set is given which leads to simple randomised rounding algorithm using linear programming formulation of the problem. In Section 4 we analyse the results obtained from the algorithm's application on a Twitter network and generated graphs and compare them for non-weighted case with the existing algorithm in [5] for -rate domination. We conclude in the Section 5.

1 Preliminaries

In this section we introduce the notation and de nitions that we use throughout this paper.

A graph or undirected graph G is an ordered pair G = (V; E) where V is a set, elements of which are called vertices or nodes, and is a set of unordered pairs of distinct vertices called edges. If G is a graph of order n, then $V(G) = fv_1; v_2; ...; v_n g$ is the set of vertices in G, d_v denotes the degree of v, and $d_v = d_v + 1$. Let N(v) denote the neighbourhood of a vertex v. Also, let N(V) = $[v_2 v N(v) \text{ and } N[V] = N(V)[V: Then <math>d_v = jN[v]$. Denote by (G) and (G)

the minimum and maximum degrees of vertices oG, respectively. Put = (G) and = (G).

A set D is called adominating set if every vertex not in D is adjacent to one or more vertices in D. The minimum cardinality of a dominating set of G is the domination number (G).

be a real number satisfying 0< 1. A set X V(G) is called Let an -dominating set of G if $jN(v) \setminus X j$ d_v for every vertex v 2 V(G) n X, i.e. v is adjacent to at least dd ve vertices of X. The minimum cardinality of an -dominating set of G is called the -domination number (G). It is easy to see that (G) (G), and (G) = (G)(G) if is su ciently close to 0. A V(G) is considered an -rate dominating set of G if for any vertex set X v 2 V(G), jN[v]\ X j d_v: The minimum cardinality of an -rate dominating set of G is called the -rate domination number (G). It is easy to see that (G) (G).

Now we consider the vertex-weighted graphs. These are nite and undirected graphs with no loops and multiple edges in which each vertex has been assigned a weight. Let w_v be the weight (cost) of each vertexv of graph G. Let $_w(G)$ denote a minimum weight of a dominating set X of G and let $_{;w}$ denote a minimum weight of a $_{v2D}$ -rate dominating set D. Finding an -rate dominating set D of G such that $_{v2D} w_v$ is minimised is the main problem studied in this paper.

2 Previous work

Variants of domination have been studied extensively and have various applications for real life problems. Smaller number of studies in domination parameters consider weighted graphs in particular.

The minimum weighted dominating set problem is one of the classi**N** P -hard optimisation problems in graph theory. Zou et al. [20] studied the minimum-weighted dominating set and the minimum-weighted connected dominating set problems on a node-weighted unit disk graph and devised approximation algorithms for these problems with performance ratios of 5 +" and 4 + " respectively. In [18] Polynomial Time Approximation Scheme (PTAS) was generalised for weighted case in polynomial growth bounded graphs with bounded degree constraint. A variant of the weighted dominating set problem - the weighted minimum independent k-domination (WMIkD) problem was studied by Yen in [19]. An algorithm linear in the number of vertices of the input graph for the WMIkD problem on trees is presented.

Discussing a more general domination set problem [3], where the direct connections are replaced with shortest paths corresponding to some measufedened on the vertices of a graph, the authors give an approximation algorithm for the vertex-weighted version. Using randomised rounding they prove the approximation ratio of $O(\log)$ for their randomised algorithm, where is the maximum cardinality of the sets of vertices that can be dominated by any single vertex, or in our case a maximum degree of the vertices in the graph. In [2], the maximum spanning star forest problem is discussed, which is the complement problem of domination set. A 071-approximation algorithm for this problem is given, and for vertex-weighted case a :64-approximation algorithm is presented.

The -domination was introduced by Dunbar et al. in [12]. Introduced by Zverovich et al.[6] the concept of -rate domination can be considered as a particular case of an dominating set in the same graph. Note that both the and -rate domination problems are known to be NP -complete. Thus it is of importance to determine bounds for and -rate domination numbers and various similar parameters. In [5] and [6] the authors explicitly provide new upper bounds and randomised algorithms for nding the and -rate domination sets in terms of a parameter and graph vertex degrees on undirected simple nite graphs by using probabilistic constructions. Their algorithm is bounded by:

(G)
$$@_1 \qquad \frac{b}{(1+b)^{1+1}} A_n;$$
 (1)

where \mathfrak{G} is a closed -degree of G and b = b (1) (c+1).

Studies of the propagation of in uence in the context of social networks carried out by Wang et al. in [16] resulted in introducing new variants of domination such as the positive in uence dominating set (PIDS) and total positive in uence dominating set (TPIDS). From the de nitions given in [16] it is easy to see that PIDS and TPIDS problems are equivalent to -dominating and -rate dominating set problems respectively for a special case when = 1=2. Wang et al. proved that both these problems areNP-hard. Thus, it is important to study approximability of the problems. In their work Dinh et al. [4] generalise PIDS and TPIDS by allowing any 0 < < 1 and show that both problems can be approximated within a factor In + O(1) and present linear time exact algorithm for trees.

3 Randomised rounding algorithm

of an integer program IP:

min

Showing our goal (4) is equivalent to showing

So we are looking for a minimum of the right hand side of (5) subject to $\int_{i=1}^{d_v} x_i = k$ (this minimum must exist by continuity and compactness). Clearly the minimum will be found when $\int_{i=1}^{d_v} x_i = k$; increasing one of thex_is without changing the others will clearly only increase the RHS. So we may assume that

$$X_{i}^{v} = k:$$

$$i=1$$
(6)

Now we can use the result from [10], Theorem 5, that shows that tail distribution function of Poisson's binomial distribution attains its minimum in binomial distribution, i.e. when all probabilities are equal. The theorem states that for two integers b, and c such that 0 b np c n, the probability P(b X c) reaches its minimum where all the probabilities $p_1 = :::= p_n = p$, unless b = 0 and c = n. Here p_i s are probabilities (or parameters) of Poisson's binomial distribution, and n and p are parameters of related binomial distribution. We apply that theorem taking two integers b and c to be our k and d_v respectively. We have that p, the equal probability is $\frac{k}{d_v}$ from (6), whence np equals our k. The theorem gives us

Thus, we will be done if we can show that

is at least $\frac{1}{2}$. Let Y be a random variable of binomial distribution with d_v trials each of probability p. Then observe that in fact $Pr(Y \ k)$ is equal to (7) above. The median of Y is bounded by bd_v pc and dd_v pe [13], but d_v p is exactly the integer k, so k is the unique median of Y. It follows from the de ning property of medians that $Pr(Y \ k) \ \frac{1}{2}$, and thus $Pr(Y \ k) < \frac{1}{2}$ and the proof is complete.

Hence, the probability is lower bounded by $\frac{1}{2}$, and the feasibility follows. Let A_i denote the event that vertex v_i is -rate dominated and let $B = \bigwedge_{i=1}^{n} A_i$ be the event that all vertices are dominated. Using a technique identical to one carried out in [3], with ampli cation approach (repeating randomised rounding $t = O(\log_2)$) times) which results in $Pr([x_i = 1]) = 1$ $(1 \quad x_i)^t$. We obtain

that the expected value of the solution resulted from randomised rounding, given that event B happens, (i.e. that the solution is feasible) is

$$E \sum_{i=1}^{\#} w_{i} x_{i} j B = \sum_{i=1}^{\#} w_{i} Pr([x_{i} = 1]jB)$$

$$= \frac{X^{n}}{\sum_{i=1}^{W_{i}} w_{i} \frac{Pr(Bj[x_{i} = 1])}{Pr[B]} Pr(x_{i} = 1)$$

$$= \frac{X^{n}}{\sum_{i=1}^{W_{i}} w_{i} Q} \frac{1}{j^{2N}(v_{i})} Pr(A_{j}) (1 (1 x_{i}))^{t}$$

$$= \frac{t}{2} \frac{X^{n}}{\sum_{i=1}^{W_{i}} w_{i} x_{i}}{O(\log_{2} OPT)}$$

Hence, there exists a particular solution that it is within ($O\log_2\;$) ratio to the optimal solution. tu

A simple randomised rounding algorithm AlgRR follows immediately, by rst solving LP and then rounding the solutions to zero or one. All vertices with ones then create an -rate domination set with the sum of the weights within $O(\log_2)$ factor of the optimal solution. We implemented AlgRR in Python.

4 Twitter UK mentions network

The Twitter data-set was collected on our behalf by Datasift, a certi ed Twitter partner, which allowed us to access the full Twitter rehose rather than being rate-limited. The data-set consists of all UK based Twitter users that sent tweets with at least one mention between 8 Dec 2011 and 4 Jan 2012 (28 days in total). Mentions are messages that include an @ followed by a username and are used to address people. Thus, if persoA posts a tweet containing \@B" that means A is addressing the tweet toB speci cally. Mentions are not private messages and can be read by anyone who searches for them. A tweet can be addressed to several users simultaneously using @ repetitively.

4.1 Data

We preprocessed the data, removing empty mentions and self-addressing which left us with 3; 614; 705 time-stamped arcs (individual mentions) from a total of 819,081 distinct usernames, or nodes. We then removed all users who didn't tweeted but just received messages, as we did not have a weight measure for them. There were approximately 50k nodes that appeared both as tweeters and receivers. We aggregated data on weekly basis and kept only two-directional arcs (thus if person A mentioned B and person B mentioned A at least once during

³ All Twitter users appearing in our data-set had selected the UK as their location.

a week there is a bi-directional edge between A and B in a weekly graph). For simplicity, we treated those bi-directional edges as undirected. This left us with 4 undirected weekly graphs with around 5k nodes in each and around:5k edges in average. For each vertex we retrieved its flout score and used it as a weight. The Klout score measures an individual's in uence based on her/his social media activity ⁴ It is a single number that represents the aggregation of multiple pieces of data about individuals' social media activity, based on a score model which is not publicly available [14]. The descriptive statistics of the Twitter mentions weekly graphs are given in Table 1 below. As the 4 mentions graphs are quite sparse, we

Table 1. Twitter mentions network statistics, ME denotes number of multi-edges in

4.2 Results

In this section we investigate how our randomised rounding algorithm AlgRR performs on some real and created networks. We also compare it with the existing -rate domination algorithm for simple (non-weighted) graphs from [5] (denoted here as AlgA). We have run both algorithms on 4 weekly Twitter random and preferential graphs. As algorithms are randomised, we have run both algorithms 100 times taking averages. Results are presented in Tables 3 and 4 below. The results show that for dense networks (networks denoted withpref-d and rnd-d) the algorithm AlgRR outperforms algorithm A signi cantly and not only on minimum weights (which would be expected, as algorithm A optimises the size of -rate dominating set, while algorithm AlgRR optimises the weights) but also on the sizes of -rate dominating sets. According to the theoretical bounds of algorithm A the probability with which each candidate vertex for -rate dominating set is selected gets close to 1 for dense network, thus

resulting in selecting all the nodes of the network. However on sparse networks such astwitt1-4 Algorithm A slightly outperforms algorithm AlgRR.

Table 3. Alpha-rate domination sets' sizes (#), weights(W) and running times(T) for AlgA, for di erent graphs and = 0.225; 0:5; 0:75 respectively.

Graph Avg#	AvgW Min# Ma	ax# Max	<w avgt(ms)<="" minw="" th=""></w>
pref-d0.25 5000	193419 5000	5000	193419 193419 12.71
pref-d0.5 5000	194267 5000	5000	194267 194267 12.87
pref-d0.75 5000	188938 5000	5000	188938 188938 13.09
rnd-d0.25 4730	182675 4689	4768	184149 180909 12.11
rnd-d0.5 4991	191934 4985	4998	192222 191573 12.35
rnd-d0.75 4998	194278 4994	5000	194343 194082 12.58
twitt10.25 4328	146960 4259	4408	149577 144171 3.21
twitt10.5 5291	179701 5222	5334	181377 176986 4.21
twitt10.75 5056	171733 4993	5113	173600 169609 3.80
twitt20.25 4612	157207 4539	4670	159495 154516 3.22
twitt20.5 5453	185872 5436	5475	186780 185343 3.77
twitt20.75 5259	179254 5217	5297	180618 177888 3.58
twitt30.25 3960	135557 3879	4031	138291 132624 2.70
twitt30.5 4897	167738 4854	4946	169551 165730 3.60
twitt30.75 4753	162770 4697	4794	164233 160732 3.24
twitt40.25 4195	142143 4119	4289	145122 139323 3.24
twitt40.5 5127	173809 5069	5183	175550 171776 3.64
twitt40.75 5036	170727 4979	5092	172826 168874 3.53

Since the algorithm AlgRR is based onLP -relaxation technique it runs in polynomial time. Our resetwOurpolynomialin Tab.no6138291LP0

Table 4. Alpha-rate domination sets' sizes, weights and running times for AlgRR, for di erent graphs and = 0.25; 0:75 respectively.

Graph	Avg#	AvgW Mir	n# Ma	x# Max	W MinW	AvgT(r	ms)
pref-d0.25	979	16647	578	1512	30845	7124	65.70
pref-d0.5	2158	51541	1498	2777	75523	28143	139.24
pref-d0.75	3489	109794	2661	4279	149581	71461	256.96
rnd-d0.25	1604	28367 9	960	2420	56342	10869	341.29
rnd-d0.5	2688	68953 ´	1843	3664	115395	34074	870.87
rnd-d0.75	3901	130250	2747	4770	180502	70105	1180.02
twitt10.25	4628	153367 4	607	4655	154355	152560	8.79
twitt10.5	4817	159901 4	792	4837	160658	158939	9.11
twitt10.75	5665	191648 5	665	5665	191648	191648	9.29
twitt20.25	4443	148065 4	413	4469	148980	147007	7.77
twitt20.5	4595	153249 4	566	4625	154311	152196	8.06
twitt20.75	5431	184481 5	5431	5431	184481	184481	8.13
twitt30.25	4191	140094 4	166	4220	141094	139172	6.99
twitt30.5	4360	145765 4	321	4391	146869	144362	7.13
twitt30.75	5159	175854 5	5159	5159	175854	175854	7.26
twitt40.25	4459	147854 4	447	4471	148289	147384	7.80
twitt40.5	4637	153552 4	624	4651	154067	153058	8.12
twitt40.75	5468	184486 5	5468	5468	184486	184486	8.16

The analysis of AlgA has shown similar spread of solutions for di erent runs, and were relatively stable. For algorithm AlgRR the results in Table 4 show signi cant di erence between minimum and maximum cardinalities of -rate dominating sets for dense networkspref-d and rnd-d. This indicates that the values of variables in the solutions obtained by LP relaxation are spread out over (0; 1) interval (i.e. are fractional). We have veri ed the spread and consistency by performing additional 200 runs for algorithm AlgRR where = 0.25 for pref-d and rnd-d networks and recordeF52 9.9ue networks and recor

-rate dominating sets does not change signi cantly compared with the results obtained from 100 runs. Thus it can be concluded that more runs are unlikely to achieve better results.



Fig. 1. Variation in size and weight for 100 runs of AlgRR, for rnd-d, = 0:5

On Figure 1 given are sizes and weights for 100 runs for AlgRR omnd-d networks where = 0.5. Similar plots were obtained for all graphs and for both algorithms.

5 Conclusion

We have explored how to pick optimal sets of individuals for interventions in social networks. If each person in network has assigned a cost, the aim was to nd a group of people having minimum sum of costs so that each individual in network has at least 100 percent of its neighbourhood in this designated group. We presented a randomised algorithm for nding approximation of minimum weight

rate domination set in graphs. We proved that this algorithm's output is within $O(\log_2$

References

- 1. N. Alon and J. Spencer. The Probabilistic Method. John Wiley, 1992.
- N. Chen, R. Engelberg, C. T. Nguyen, P. Raghavendra, A. Rudra, and G. Singh. Improved approximation algorithms for the spanning star forest problem. In Proc. APPROX 2007, LNCS, pages 44{58, 2007.
- 3. N. Chen, J. Meng, J. Rong, and H. Zhu. Approximation for dominating set problem with measure functions. Computing and Informatics, 23:37{49, 2004.
- 4. T. N. Dinh, Y. Shen, D. T. Nguyen, and M. T. Thai. On the approximability of positive in uence dominating set in social networks. Journal of Combinatorial Optimization, pages 1{17, 2012.
- A. Gagarin, A. Poghosyan, and V. Zverovich. Randomized algorithms and upper bounds for multiple domination in graphs and networks. Discrete Applied Mathematics, 161(4-5):604 { 611, 2013.
- A. Gagarin, A. Poghosyan, and V. E. Zverovich. Upper bounds for alphadomination parameters. Graphs and Combinatorics, 25(4):513(520, 2009.
- 7. M. R. Garey and D. S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York, NY, USA, 1979.
- C. Greaves, P. Reddy, and K. Sheppard. Supporting behaviour change for diabetes prevention. In P. Schwarz, P. Reddy, C. Greaves, J. Dunbar, and S. J., editors, Diabetes Prevention in Practice., pages 19{29. Dresden: Tumaini Institute for Prevention Management, 2010.
- A. Hagberg, D. Schult, and P. Swart. Exploring network structure, dynamics, and function using networkx. In Proceedings of the 7th Python in Science Conference (SciPy2008), Passadena, CA USA, pages 11{15, August 2008.
- 10. W. Hoe ding. On the distribution of the number of successes in independent trials. The Annals of Mathematical Statistics , 27(3):713{721, 1956.
- 11. P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. Phys. Rev. E, 65:026107, 2002.
- R. L. J.E. Dunbar, D.G. Ho man and L. Markus. -domination. Discrete Math., 211:11{26, 2000.
- 13. R. Kaas and J. Buhrman. Mean, median and mode in binomial distributions. Statistica Neerlandica, 34(1):13{18, 1980.
- 14. Klout. "http://klout.com/corp/klout _score".
- 15. T. Valente. Network interventions. Science, 337(6090), 2012.
- F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan. On positive in uence dominating sets in social networks. Theoretical Computer Science, 412(3):265 { 269, 2011.
- 17. Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In In SRDS, pages 25(34, 2003.
- Z. Wang, W. Wang, J. Kim, B. M. Thuraisingham, and W. Wu. Ptas for the minimum weighted dominating set in growth bounded graphs. J. Global Optimization , 54(3):641{648, 2012.
- 19. W. C.-K. Yen. Algorithmic results of independent k-domination on weighted graphs. Chiang Mai Journal of Science, 38:58{70, 2011.
- F. Zou, Y. Wang, X.-H. Xu, X. Li, H. Du, P. Wan, and W. Wu. New approximations for minimum-weighted dominating sets and minimum-weighted connected dominating sets on unit disk graphs. Theoretical Computer Science,