

Small Worlds by Design

Donald P. Gaver,^{*} Patricia A. Jacobs^{*} and Lyn C. Thomas[†] FIMA

Abstract

Connecting all the members of a large group involves many exchanges; think of it as clinking glasses at a social gathering. Allowing second-hand exchanges/clinks where the group splits into subgroups who clink glasses, and then representatives from each subgroup clink with each other – second-hand clinking – can decrease the number of clinks. This paper finds the optimal number of subgroups necessary to minimise the total number of clinks. It also looks at what is optimal if there is third and higher level clinking. Examples are given of such problems arising in parallel computer systems, social media, communication systems, and sports tournaments.

1 Introduction

Social events, such as parties, often promote interpersonal contact by an initial round of mutual glass-clinking. One clink between each individual of a group of n means $n(n-1)/2$ clinks in all, quite feasible in small settings, but impossible in large ($n \approx 50$ or so).

In Figure 1,

$$C_1(n) = \frac{n(n-1)}{2}$$

$$C_1(12) = 66 \text{ total connections.}$$

One way to accomplish such contact is by accepting *second-hand clinking*: split the whole group of n into $n(1)$ roughly equal size subgroups; promote first-hand clinking within each

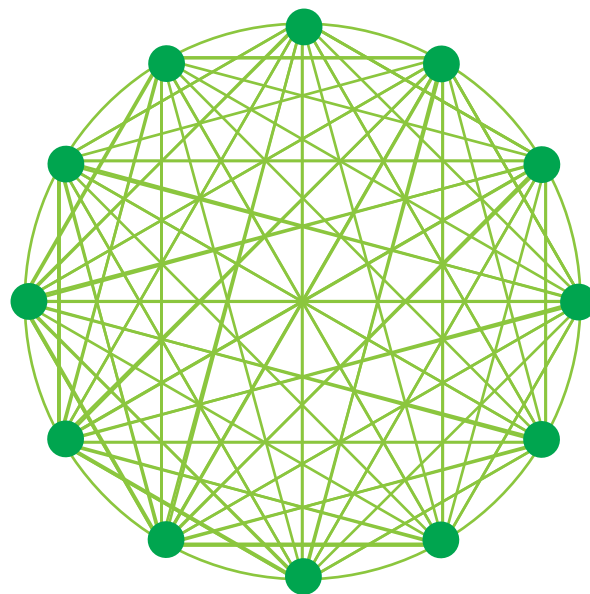


Figure 1: Single community (one world), $n = 12$.

subgroup. So $n(1)$ denotes the number of these first-level subgroups while there are n individuals in total. Select a single member of each first-level subgroup to represent that subgroup in another small, representative, subgroup; and let the representatives in the latter group exchange clinks. In this way each individual clinks/connects with each other individual with a much smaller maximum number of clinks, $C_2(n, n(1))$. What is the optimal size of the first-hand subgroups/communities $\lfloor n/n(1) \rfloor$ or $\lfloor n/n(1) \rfloor + 1$ (where $\lfloor x \rfloor$ is the largest integer less than or equal



to x) so as to minimise the total number of clinks/connections? This is, in effect, designing a *small world*, a network with a small number of connections between any two individuals. For example one group of 12 people would result in 66 clinks; splitting the 12 people into four communities with three people each results in $C_2(12, 4) = 4 \times 3 + 6 = 18$ total connections. This compares with $C_2(12, 3) = 15$, $C_2(12, 2) = 31$ and $C_1(12) = 66$. This illustrates the beneficial effect of wise group/subgroup design.

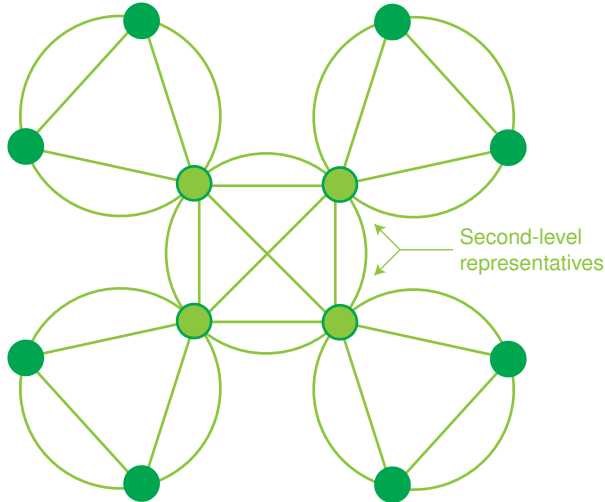


Figure 2: 12 members with second-hand clinking.

In Figure 2, $C_2(12, 4) = 4 \times 3 + 6 = 18$ total connections with four groups of three people each.

Before addressing the above simplistic question, observe that the general situation described represents various operational scenarios involving connections and communications. One example occurs when deciding how to connect central processing units (CPUs) in a parallel processor computer. Is it necessary to connect all units together or can we collect them in clusters and then connect between the representatives of the different clusters?

On a more social level, how many friends should a group on Facebook have so as to be able to spread messages throughout the group? All members of the group are split into subgroups where they make friends with everyone in the subgroup. One from each group then is made the representative for the group and the representatives become friends with each other.

A different situation in which a similar problem occurs is the setting of ATM interchange fees by banks in the same ATM network. Each year, banks meet to set fees for ATM transactions when customers of one bank uses a different bank's ATM. Meetings to set the fee can be decreased by having groups of banks deciding on their individual interchange fees and then electing a representative to meet with representatives from other groups. Another example occurs in organisations which have local branches, regional groups and a national group. The local group 'clinks glasses' by discussing and deciding matters. They elect representatives for the next level who convey the views of the local group and so on upwards. Examples range from church organisations to political parties. Finally it is also the structure of many sporting tournaments where the clinking is when teams play each other and the winner of the group goes up a level to play other winners. So in the end the winner will have 'beaten' every other team if only sometimes at second or third hand. We will discuss later the more common variant of this where two representatives – the winner and runner-up of the group go forward to the next level.

2 Optimal two-tier grouping

Divide n individuals into $n(1)$ subgroups of (approximately) equal size. Each member of each subgroup then connects ('clinks') once with each fellow subgroup member. Then a single member representative of each subgroup joins a higher level group and all members of that group connect ('clink'). What is the (approximate) number, $n^\#(1)$, that minimises the total number of connections? Putting $C_2(n, n(1))$ for the total number of connections/clinks as described above,

$$C_2(n, n(1)) \cong \underbrace{n(1) \left[\left(\frac{n}{n(1)} \right) \left(\frac{n}{n(1)} - 1 \right) \right] \frac{1}{2}}_{\text{Within}} + \underbrace{n(1)(n(1) - 1) \frac{1}{2}}_{\text{Between}}. \quad (1)$$

Approximate by neglecting the distinction between $n/n(1) - 1$ and $n/n(1)$ and between $n(1)$ and $n(1) - 1$:

$$C_2^\#(n, n(1)) \approx \frac{1}{2} \left[\frac{n^2}{n(1)} + n^2(1) \right]. \quad (2)$$

Differentiate (2) on $n(1)$:

$$\frac{d}{dn(1)} C_2^\#(n, n(1)) = -\frac{1}{2} \frac{n^2}{(n(1))^2} + n(1). \quad (3)$$

A second differentiation verifies convexity, so the (near) minimising value of the primary subgroup number is

$$n^\#(1) = \frac{n^{2/3}}{2^{1/3}} \equiv L_2 n^{p_0(2)}. \quad (4)$$

Although (2) varies from (1) by $[n + n(1)]/2$, the $n(1)$ that maximises (1) is very similar to (4) when n is less than 100. Of course the value of $n^\#(1)$ is not likely to be an integer that divides the total, n , into equal-sized integer-valued subcommunities, so we adjust to whatever of the nearest integers give the minimum value. Numerical exploration for limited n (e.g., $n = 50$) shows that such a procedure gives very nearly the minimum-connection allocation. The approximate optimal size of each of the first-tier communities is

$$g_2(1) = \frac{n}{\left(\frac{n^{2/3}}{2^{1/3}} \right)} = (2n)^{1/3}. \quad (5)$$

With the two-stage grouping in place, the total number of connections/clinks, both first and second level, is approximately

$$C_2^\#(n, n^\#(1)) = \frac{3}{2^{5/3}} n^{4/3} = \frac{3}{2} n^\#(1)^2 \equiv K_2 n^{p_C(2)}. \quad (6)$$

Note that

$$K_2 = \frac{1}{2} \left[\frac{1}{L_2} + L_2^2 \right]. \quad (7)$$

We take R to be the actual minimum using integer numbers in each subgroup. Insistence on direct connections/clinks between all participants results in the number of connections being about $C_1^\#(n) = n^2/2$; so by allowing second-level connection (through a representative) to count the system requires a fraction F less clinking where

$$F = \frac{C_2^\#(n, n^\#(1))}{C_1^\#(n)} \approx \frac{\frac{3}{2^{5/3}} n^{4/3}}{n^2/2} = 3 \frac{1}{2^{2/3} n^{2/3}} = \frac{3}{2} \frac{1}{n^\#(1)}. \quad (8)$$

* Naval Postgraduate School, Monterey, California

† University of Southampton

Table 1 displays the number of clinks required for various subgroups of a total population allowing for the adjustment to get integer values. If $n = 50$ only about 12% as many connections are required in the two-level small world as in the one large world case.

3 General recursion

The idea of second-hand clinking can be extended to third-hand and higher tiers/levels of clinking. In this, a representative of a subgroup at a lower level meets representatives of some of the other such subgroups to form a subgroup at the next higher level. If one allows third-level (third-hand) clinking, then define $n_3(1)$ to be the number of first-level subgroups each of $\lfloor n/n_3(1) \rfloor$ or $(\lfloor n/n_3(1) \rfloor + 1)$ individuals. At the second level there are $n_3(2)$ groups each of $\lfloor n_3(1)/n_3(2) \rfloor$ or $(\lfloor n_3(1)/n_3(2) \rfloor + 1)$ representatives from the lower tier. Finally at the third tier there is one group consisting of $n_3(2)$ members each of which represents one of the second tier subgroups. For example, one possibility for 12 people is as follows: the first tier contains six groups of two people each; the second tier consists of three groups of two people each; and the third tier has one group of three people. This is shown in Figure 3 with three levels of clinking, where first-level clinking is thin lines, second-level clinking is the dotted lines and third level clinking is the thicker lines.

If $C_3(n, n_3(1), n_3(2))$ represents the number of clinks under three tiers one gets the recursion:

$$C_3(n, n_3(1), n_3(2)) = \frac{1}{2} \frac{n^2}{n_3(1)} + C_2(n_3(1), n_3(2)) \quad (9)$$

$$C_3\left(n, n_3(1), n_3^\#(2)\right) = \frac{1}{2} \frac{n^2}{n_3(1)} + n_3(1)^{p_C(2)} K_2 \quad (10)$$

where (6) is substituted into (9) with $n_3(1)$ and $n_3(2)$ replacing n and $n(1)$. This gives $K_2 = 3/(2^{5/3})$ and $p_C(2) = 4/3$; the form is inferred from (6) and (7). Next differentiate (10) and set the derivative equal to 0 to find optimum numbers of subgroups at the first tier:

$$\begin{aligned} n_3^\#(1) &= \frac{1}{2^{4/7}} n^{6/7} \\ &= \left[\frac{1}{2p_C(2)K_2} \right]^{1/[1+p_C(2)]} n^{2/[1+p_C(2)]} \\ &\equiv L_3 n^{p_0(3)}. \end{aligned} \quad (11)$$

From (4)

$$n_3^\#(2) = L_2 n_3^\#(1)^{p_0(2)}. \quad (12)$$

Combining (11) and (10) results in

$$\begin{aligned} C_3^\#(n, n_3^\#(1), n_3^\#(2)) &= \frac{1}{2} \frac{n^2}{L_3 n^{2/[1+p_C(2)]}} + \left[L_3 n^{1/[1+p_C(2)]} \right]^{p_C(2)} K_2 \\ &= \left[\frac{1}{2L_3} + L_3^{p_C(2)} K_2 \right] n^{2p_C(2)/[1+p_C(2)]} \\ &\equiv K_3 n^{p_C(3)}. \end{aligned} \quad (13)$$

Further exploration shows that a power-law recursion holds for more tiers.

Let $n_m^\#(i)$ be the optimal number of individuals in the level $i = 1, 2, \dots, m-1$ subgroup when there are m tiers of connection. Let $C_m^\#$ be the minimising total number of connections

under the same conditions. A recursive argument on m yields

$$C_{m+1}^\# = n^{p_C(m+1)} K_{m+1} \quad (14)$$

$$n_{m+1}^\#(1) = L_{m+1} n^{p_0(m+1)} \quad (15)$$

where

$$p_0(m+1) = \frac{2}{p_C(m)+1} \quad (16)$$

$$p_C(m+1) = \frac{2p_C(m)}{p_C(m)+1}$$

and

$$L_{m+1} = \left[\frac{1}{2p_C(m)K_m} \right]^{1/(p_C(m)+1)} \quad (17)$$

$$K_{m+1} = \frac{1}{2L_{m+1}} + L_{m+1}^{p_C(m)} K_m. \quad (18)$$

The optimal numbers of individuals in each of the other subgroup tiers are

$$\begin{aligned} n_{m+1}^\#(2) &= L_m n_{m+1}^\#(1)^{p_0(m)} \\ n_{m+1}^\#(3) &= L_{m-1} n_{m+1}^\#(2)^{p_0(m-1)} \\ &\vdots \\ n_{m+1}^\#(k) &= L_{m+2-k} n_{m+1}^\#(k-1)^{p_0(m+2-k)} \end{aligned} \quad (19)$$

for $k = 2, 3, \dots, m-1$.

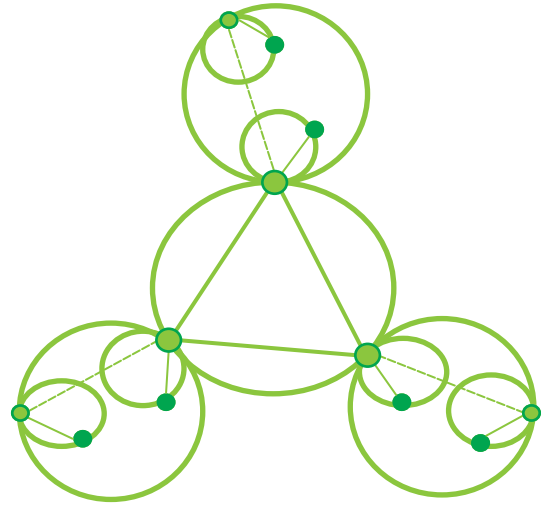


Figure 3: 12 members with third-hand clinking.

Applying these results to third-hand clinking gives the results in Table 2. So the approximations for the number of groups works well and three tiers decreases the needed clinking/connections.

4 Two first-level representatives

We mentioned earlier that in sporting tournaments there are often two representatives who are chosen from the first-level groups. Each representative from a group is put in a different group at the next level. Call them Groups A and B. One can then apply normal or higher level clinking to these groups. This ensures that two members of the original group have two ways in which their clinking can occur – one through the representatives in Group A and one through the representatives in Group B. In fact if there is a higher level of clinking with one representative from Group A and one from Group B having a final clink (just like the ‘final’ of a sporting competition) there would be four ways any two

people in the original group have clinked via high order clinking. The first is through the Group A representatives of each person; the second through the Group B representatives of each person clinking and two ways through the Group A representative of one person clinks with the Group B representative of the other through that final higher level clink. The ideal number of first-level groups can be estimated using an adaption of (9). If there are $n(1)$ first-level groups and we take $C_2(n(1))$ to be the number of subsequent higher level clinks involved if there are $n(1)$ first-level groups, then instead of the total number of clinks (excluding the final clink between one representative from A and one representative from B) being

$$C_3(n, n(1)) = \frac{1}{2} \frac{n^2}{n(1)} + C_2(n(1)), \quad (20)$$

which is a generalisation of (9), it would be

$$D_3(n, n(1)) = \frac{1}{2} \frac{n^2}{n(1)} + 2C_2(n(1)) \quad (21)$$

since there are $C_2(n(1))$ higher order clinks in Group A and a similar number in Group B. In the simplest case where there is only second-hand clinking, this gives rise to the optimal $n(1)$ being $n(1) = (n/2)^{2/3}$. Applying this to the cases with $n = 10, 50, 100$ and $1,000$ gives Table 3 where the total number of clinks ignores the final single clink.

Of course in sports tournaments the objective is not to minimise the number of matches played but instead play the number

of matches required by the broadcasting rights. However, this approach can allow one to look at the tournament structure that provides the requisite number of clinks/matches/connections.

5 Discussion

The focus of this note is on network design to optimise a particular measure of performance. Ours is an *architectural* study, differing from the great majority of previous works, for example [1, 2]. These aim to describe plausible *behavioural* dynamics that lead to small-world connections. Behavioural models apparently stem from classical experiments [3]. Our formulations for increasing tiers eventually lead to an optimised *Moore graph* [4]. Changes in the system performance metric will naturally lead to altered graph configurations. For instance, recognition of link or node unreliability, or congestion at sensor-servers, will demand combined architectural and sensor behavioural dynamic accommodation requiring further study.

REFERENCES

- 1 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small world' networks, *Nature*, vol. 393, pp. 330–442.
- 2 Newman, M.E.J. (2002) The structure and function of complex networks, *SIAM Rev.*, vol. 74, pp. 47–97.
- 3 Travers, J. and Milgram, S. (1969) An experimental study of the small world problem, *Sociometry*, vol. 32, pp. 425–442.
- 4 Watts, D.J. (1999) *Small Worlds: The Dynamics of Networks Between Order and Randomness*, Princeton Studies in Complexity, Princeton University Press.

Table 1: Second-level clinking examples

Total number of agents, n	10	50	100	1,000
Number of connections with one community	45	1,225	4,950	499,500
Optimal number of subgroups, $n(1)$ (approximation from (4))	4 (3.6)	11 (10.8)	17 (17.1)	79 (79.37)
R , total number of connections	14	145	381	8,919
F^\ddagger	31.1%	11.8%	7.7%	1.8%

Table 2: Third-level clinking

Total number of agents, n	10	50	100	1,000
Number of clinks with one level	45	1,225	4,950	499,500
Approximate $n(1)$ (from (11))	4.8	19.2	34.9	250.8
Approximate $n(2)$ (from (12))	2.3	5.7	8.5	31.6
$n(1)$ for 3-level clinking	5	19	35	250
$i(1)$: $i(1)$ or $i(1) + 1$ in first-level group	2	2	2	4
Number of first-level (1) clinks	5	43	95	1,500
$n(2)$ for 3-level clinking	2	6	9	31
$i(2)$: $i(2)$ or $i(2) + 1$ in second-level group	2	4	3	8
Number of second-level (2) clinks	4	21	51	884
Number of third-level (3) clinks	1	15	36	465
R , total number of clinks	10	79	182	2,849
F^\ddagger	22.2%	6.5%	3.7%	0.6%

Table 3: Optimal grouping with two first-level representatives

Total number of agents, n	10	50	100	1,000
Number of clinks/connections with one community	45	1,225	4,950	499,500
Optimal number of groups, $n(1)$ (approximation $(n/2)^{2/3}$)	3 (3.68)	9 (8.5)	14 (13.6)	63 (63.0)
R , total number of connections	18	187	490	11,346
F^\ddagger	40.0%	15.3%	9.9%	2.3%

‡ ratio of number of clinks if second level allowed compared with only first level: $R/(\text{number of connections with one community})$