EQUINE VETERINARY JOURNAL
*Equine vet. J.* (2003) **35** (6) 586-589

# Small-world topology of UK racing: the potential for rapid spread of infectious agents

**R. M. CHRISTLEY*** and **N. P. FRENCH**

*Department of Veterinary Clinical Science, University of Liverpool, Leahurst, Neston, Wirral CH64 7TE, UK*

**Keywords:** horse; infectious disease epidemiology; transmission; contact network

## Summary

*Reasons for performing study:* **The topology of the network of contacts between individuals has important effects on infectious disease dynamics within a population. Here we examine for the first time a network of contacts between training yards that occurred through racing.**

*Objectives:* **To explore the topology of this network and to consider the effects of the network on the potential for disease transmission.**

*Methods:* **Race data from one week was analysed. Contacts were defined as occurring between trainers that raced horses in the same race and hence one trainer could contact another trainer several times. A connection was said to exist between trainers who contacted each other at least once. The network of contacts and connections that occurred during the study period was reconstructed and analysed.**

*Results:* **All 466 trainers formed a single large network. The network of contacts had a short average path length and high clustering and was, therefore, characteristic of a 'small world network'. The probability distribution of the number of contacts was scale-free, whereas that for the number of connections followed a single-scale. The effect of the network would be to increase $R_0$, such that an agent that would tend toward extinction in a homogenously mixing population may persist in the observed network.**

*Conclusions:* **The observed small world network topology has important implication for the transmission and, therefore, the control of infectious agents in this population.**

*Potential relevance:* **Effective disease control and surveillance must take account of the contact structure of the population. Further studies investigating other contact definitions and other populations are now required.**

## Introduction

An essential feature of infectious diseases is their transmission from host to host, in many instances requiring close contact between individuals. The pattern of contacts ('contact network') within a population has therefore a major influence on spread. However, the effect of contact networks on transmission dynamics has been limited mostly to transmission of sexually transmitted diseases of man (Ghani *et al.* 1997; Liljeros *et al.* 2001).

Importantly, transmission is affected by the topology of a network, rather than solely by the attributes of the individuals in the network. A characteristic feature of many real networks is the co-existence of clustering and short path lengths (Watts and Strogatz 1998). Clustering is the tendency for the formation of cliques, where nodes that contact a common node also contact each other. A node is the unit of study (e.g. individual animals or racehorse trainers or, more specifically, training yards). The path length is the shortest distance (via other nodes) between any 2 nodes in the network; length of 1 indicates nodes contact each other directly, length of 3 that there are 2 intervening nodes. Because of the relatively short distance between nodes, these networks are referred to as 'small-worlds' (Watts and Strogatz 1998). The existence of a small-world network structure is confirmed by comparison of the observed network with an equivalent random network, where the same number of nodes form the same number of connections, made at random. Small-world networks have a similar average path length, but much greater clustering, compared to a random network. The characteristics of the contact network have important effects on the transmission of infectious diseases within populations (Lloyd and May 2001); and on the efficacy of vaccination programmes (Zanette and Kuperman 2002). In addition, understanding of the network topology may enable development of novel disease control or surveillance schemes (Liljeros *et al.* 2001; Lloyd and May 2001).

A prominent feature of many epidemics, e.g. UK foot-and-mouth disease (Ferguson *et al.* 2001) and South African equine influenza outbreaks (Guthrie *et al.* 1999), is the rapid spread of agents across long distances, suggesting the important role of contact networks during outbreaks of animal diseases. Importantly, racing brings together horses from geographically remote regions and factors associated with racing, including stress, which may facilitate disease transmission (Wong *et al.* 1992). The aim of our study was to investigate the topology of one type of contact in racing in the UK, in order to suggest the potential for spread of infectious agents.

## Materials and methods

Data were obtained from Computer Raceform[1], a database of all horses racing in the UK. All races occurring between October 1st and 7th 2001 were selected and downloaded to a spreadsheet

---

*Author to whom correspondence should be addressed.

*[Paper received for publication 08.01.03; Accepted 29.04.03]*

programme[2] . This week was selected from the most recently available data (2001) as a week when both National Hunt and Flat racing occurred. The data for the 7 day study period were analysed using the Payek Programme for Large Network Analysis v0.85 (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

In this study, the unit of interest were UK racehorse trainers (nodes). Nodes form networks by link with each other by 'edges' or 'arcs'. Edges are undirected contacts, whereas arcs are directed contacts. In this study, all contacts were considered to be undirected, with the edges being horses that competed in the same race. A connection between trainers was defined as occurring when they raced together at least once. The number of connections made by one trainer with other trainers is $k_{trainer}$. A contact occurred every time a trainer races a horse against the horse of another trainer. The number of contacts a trainer has is $k_{total}$. If trainer A entered different horses in each of 2 races, direct contact with trainers of all horses in those races (i.e with a path length (L) of 1) would occur; and each trainer in one race would contact each trainer in the other race via trainer A (i.e with L of 2).

The degree ($k_v$) for trainer v was defined as the number of edges (connections) linking to that trainer (i.e. the number of other trainers which trainer v competed against; the neighbours of trainer v). Average degree ($\langle k \rangle$) for all trainers was calculated (total number of edges/number of nodes). The clustering coefficient (CC) was calculated as the average of the fraction of allowable edges between neighbours of trainer v that actually occurred. That is:

$$CC = \frac{\Sigma C_v}{n}$$

where n is the number of trainers and $C_v$ was calculated as:

$$C_v = \frac{2 \times \Sigma G_v}{k_v (k_v - 1)}$$

where trainer v has $k_v$ edges and $\Sigma G_v$ is the sum of all edges between trainers that contact trainer v (Watts and Strogatz 1998).

A random network with an equivalent number of nodes (466) and average degree ($\langle k \rangle$) was constructed and the $CC_{rand}$ and $L_{rand}$ calculated similarly.

Although the presence of a connection between 2 trainers was a binary phenomenon (either a connection was present or it was not), multiple individual contacts could occur. The frequency distribution of the number of contacts and connections and of the number of contacts per connection, was examined.

The effect of heterogeneous mixing on the basic reproductive ratio ($R_0$) was calculated as:

$$R_0 = \rho_0 [1 + (CV)^2]$$

where $\rho_0$ is the average number of secondary infections produced by an infected individual in a wholly susceptible population, if the population mixed homogenously (May *et al.* 2001) and CV is the coefficient of variation of the node connectivity, calculated as:

$$CV = \frac{\sigma}{m} = \sqrt{\frac{[k_{trainer}^2]}{[k_{trainer}]^2} - 1}$$

where σ is the standard deviation and m the average number, of edges per trainer, $\langle k_{trainer} \rangle$ is the average number of connections made by each trainer (i.e. $m = \langle k_{trainer} \rangle / 2$, as each edge is shared by 2 trainers) and $\langle k_{trainer}^2 \rangle$ is the average of the square of the number of connections made by each trainer (May *et al.* 2001; May and Lloyd 2001).

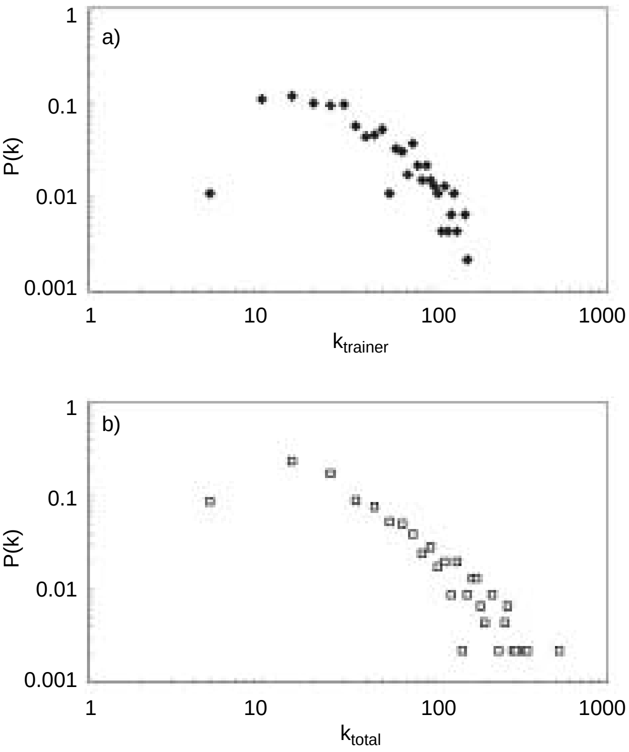

*Fig 1: Distribution of connections and contacts for 466 racehorse trainers that participated in racing between October 1st and 7th 2001. a) Probability distribution of the number of connections between trainers. b) Probability distribution of the number of contacts between trainers. The distribution of contacts follows a single scale (exponential) distribution. The distribution of the number of connections appears close to linear (particularly between k = 20 and k = 120) suggesting a scale-free distribution.*

## Results

During the 7 days race data analysed there were 149 races, held at 18 tracks. Racing occurred on more than one day at only one racetrack (4 days). During all of October 2001, 632 trainers raced one or more horses. Of these, 466 (74%) raced at least once during the study period; 200 raced only in flat races, 183 only in National Hunt races and 83 in both flat and National Hunt races.

The trainers formed a single network that was characteristic of a small-world network. Mean L was 2.26, only slightly greater than that of a random network with the same number of nodes and average k (1.93). Maximum L between 2 trainers was 4. In addition, clustering coefficient (0.66) was substantially greater than that for the random network (0.09). The average number of connections per trainer ($\langle k_{trainer} \rangle$) was 41 (s.d. 33).

The log-log plot of the probability distribution of connections ($k_{trainer}$) suggests an exponential decay of the distribution (Fig 1). When all contacts ($k_{total}$) were considered, the degree distribution more closely approximates a scale-free distribution (i.e. linear on a log-log plot), particularly for moderate k (between approximately 20 and 120). The relationship between $k_{trainer}$ and

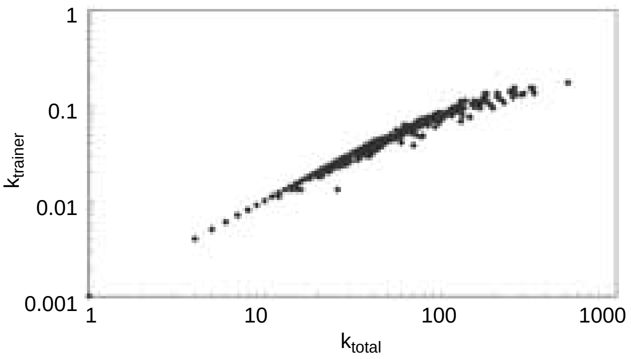

*Fig 2: Association between the number of contacts ($k_{total}$) and the number of connections ($k_{trainer}$) for 466 racehorse trainers that participated in racing between 1st and 7th October 2001.*

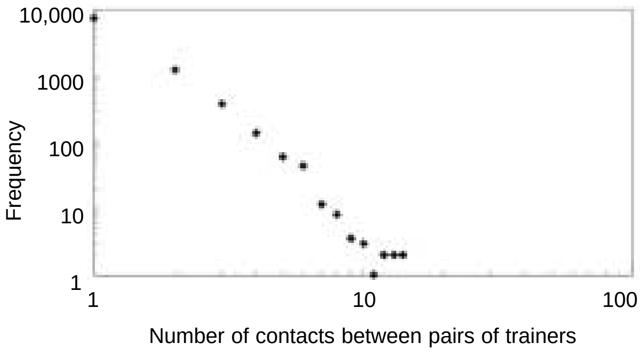

*Fig 3: The distribution of the number of contacts between individual pairs of racehorse trainers that participated in racing between October 1st and 7th 2001. For example, almost 10,000 pairs of trainers contacted each other only once and approximately 1000 pairs contacted each other twice, during the week of study.*

$k_{total}$ suggests there was little increase in $k_{trainer}$ after approximately 100 connections, despite further increases in $k_{total}$ (Fig 2). The majority of connections between trainers formed with only one contact between those trainers (Fig 3).

The effect of the network structure on the basic reproductive ratio was assessed. An agent, which in a homogenous population had $\rho_0 = 0.6$, would have $R_0 = 1$ in this network of trainers. In addition, the short average path length (and maximum path length) indicate that, were a disease agent to be introduced into any one training yard, all other trainers in the network would be only a few steps from that yard.

## Discussion

This appears to be the first study to apply network theory to the analysis of the contact structure of an equine population. The methods used have great potential in the development of understanding of infectious disease transmission and control in populations. The observed network of contacts during racing was a small-world network. Further work is required to explore the topology of contact networks using different contact definitions. However, given the ubiquity of small-world architecture in natural and man-made networks, we suggest that the basic characteristics of these networks would be similar (Watts and Strogatz 1998; Amaral *et al.* 2000; Liljeros *et al.* 2001).

The most appropriate definition of contact differs between infectious agents. We defined a contact as occurring each time a horse from one training yard raced against a horse from another training yard. Hence a trainer may have multiple contacts with another trainer. Two trainers were connected in the network if they contacted each other at least once during the study period. These definitions may underestimate the actual connections and contacts between trainers at racecourses, if mechanisms such as shared equipment or contact between personnel are important aspects of disease transmission. Conversely, horses with transmissible infectious disease may be less likely to participate in racing.

The network of trainers observed in this study had characteristics of a small-world network; short path lengths and clustering (Watts and Strogatz 1998). The maximum path length between any 2 trainers was small and on average was just over 2. That is, on average, a trainer linked to any other trainer via one intermediate trainer. This was despite UK racing being grouped into flat and National Hunt racing. The short path length between trainers engaged solely in flat racing and those involved only in NH racing must act through those trainers participating in both. Hence, despite the structure of racing in the UK, the contacts occurring over 7 days resulted in formation of a single network in which three-quarters of all the trainers that raced in October 2001 had the potential to be exposed to infectious agents though a small number of links.

There is likely to be a limit to the number of contacts trainers can make in a short period, related to the number of horses under their care and to the number of races during that period. For example, the trainer with the highest degree entered 36 races during the study period (out of a possible 106 flat races), that is 1 in 3 available races, averaging over 5 races/day. This may be close to the upper practical limit, given constraints of staffing and transportation.

Inclusion of all contacts (as apposed to connections) results in a better approximation of a scale-free distribution, suggesting that more highly connected trainers duplicated existing connections to other more highly connected trainers. That is, the more highly connected trainers 'waste' edges contacting trainers to whom they are already connected. For example, the most connected trainer directly contacted 61% of all flat race trainers. Inevitably, this trainer tended to contact other well-connected trainers (i.e. those with high $k_{trainer}$; median = 61, inter-quartile range 38–94), whereas those not contacted tended to be those with low $k_{trainer}$ (median 20, IQR 13–32, Mann-Whitney Test $P<0.0001$). Even by entering more races, many of the contacts made by highly connected trainers are likely to be reformed contacts and only a limited number will be new connections. Hence, there is evidence of assortative (like-with-like) mixing in this network (i.e. highly connected trainers tend to race against other highly connected trainers).

Although the network described accounted for only a short period, somewhat similar networks would occur continually throughout the year. Selection of different periods could have resulted in solely flat or National Hunt racing being observed, or in inclusion of major race meetings. This would alter the network, but we hypothesise that the resulting network structure would be broadly similar and constitute a small-world network. Increasing the time span studied would asymptotically increase the number of nodes and the number of contacts and connections between nodes. However, the time span of the network was appropriate given the short-lived nature of the contacts and the expected duration of infectiousness of many of the diseases likely to affect racing horses.

The small-world characteristics have important implications for the potential for spread of infectious diseases and for disease control (May and Lloyd 2001). The network structure described here would affect the reproductive ratio of infectious agents. In our simplified network, an infectious agent that would die out in a homogeneously mixing population, say with $R_0 = 0.6$, would have an $R_0$ of 1 in the network described and hence could remain endemic. As $R_0 \propto \beta D$ (where $\beta$, is the transmission parameter and D is the disease duration), an agent may remain endemic despite reduced transmission parameter and/or disease duration. Glass *et al.* (2002) reported that vaccination against Equine Influenza virus reduced $\beta$, from 1.9 to 0.5 and D from 5.5 days to 1.5 days, which led to a $R_0$ decreasing below the threshold for endemicity (0.7, compared to 10.2 in an unvaccinated population). Although this model was of transmission within and not, as in our study, between yards and our study treated contacts as permanent rather than dynamic events, the results suggest that heterogeneous mixing may enable equine influenza to invade the UK racehorse population, despite control measures and without the reintroduction of virus from outside (e.g. unvaccinated young, breeding or pleasure horses).

The majority of connections in the network of racehorse trainers are formed by a single contact. This has important implications for the interpretation of the effect of the network structure on disease transmission. The fixed, rather than dynamic, nature of the described network was a major simplification of reality. When disease transmission is considered in real time, trainers could only transmit agents to others after they acquire the agent. Ignoring this shortens the average path length observed during the study period. However, as racing occurs on most days of the year, we would expect a somewhat similar network to occur in most 7 day periods. Models that investigate the real time dynamics of infectious agents in networks are being developed, but have not yet been used widely or evaluated.

The small-world structure of UK racing has implications for infectious disease management. Scale-free small-world networks are, in general, relatively immune to random failure of nodes, but are very susceptible to attack on the most connected nodes (Albert *et al.* 2000). That is, random disease control programmes (such as voluntary vaccination) may have little overall impact, yet a programme targeting the yards of the most highly connected trainers may be effective, even if only a minority of trainers were involved in the programme. Hence, targeting highly connected trainers for disease prevention and surveillance may be one approach to disease management. However, networks with exponential degree distributions and assortative networks are more robust to such control methods than their disassortative counterparts (Newman 2002), suggesting that many trainers would need to be targeted.

This study greatly simplifies the potential network of disease transmission between race yards; contact is also likely to occur at racetracks outside a particular race and on training gallops or via shared vehicles. In addition, agents that survive for some time away from the host may not require direct contact at all. Clearly there is a need to quantify accurately the networks for specific important equine diseases, based on appropriate contact definitions, both at the racecourse and elsewhere. However, the current investigation suggests that the networks observed in such studies are likely to be small-world in character and to have important implications on disease transmission.

## Manufacturers' addresses

[1] Raceform Ltd, Newbury, Berkshire, UK.
[2] Microsoft Excel 2000, Microsoft Corporation, Redmont, Washington, USA.

## References

Albert, R., Jeong, H. and Barabasi, A.-L. (2000) Error and attack tolerance of complex networks. *Nature* **406**, 378-382.

Amaral, L.A.N., Scala, A., Barthélémy, M. and Stanley, H.E. (2000) Classes of small-world networks. *PNAS* **97**, 11149-11152.

Ferguson, N.M., CDonnelly, C.A. and Anderson, R.M. (2001) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact on interventions. *Science* **292**, 1155-1160.

Ghani, A.C., Swinton, J. and Garnett, G.P. (1997) The role of sexual partnership networks in the epidemiology of Gonorrhea. *Sex. Trans. Dis.* **24**, 45-56.

Glass, K., Wood, J.L.N., Mumford, J.A., Jesset, D. and Grenfell, B.T. (2002) Modelling equine infuenza 1: a stochastic model of with-yard epidemic. *Epidemiol. Infect.* **128**, 491-502.

Guthrie, A.J., Stevens, K.B. and Bosman, P.P. (1999) The circumstances surrounding the outbreak and spread of equine influenza in South Africa. *Rev. sci. tech. Off. int. Epiz.* **18**, 179-185.

Liljeros, F., Edling, C., Amaral, L.A.N., Stanley, H.E. and Åberg, Y. (2001) The web of human sexual contacts. *Nature* **411**, 907-908.

Lloyd, A.L. and May, R.M. (2001) How viruses spread among computers and people. *Science* **292**, 1316-1317.

May, R.M. and Lloyd, A.L. (2001) Infection dynamics on scale-free networks. *Phys. Rev. E.* **64**, 066112.

May, R.M., Gupta, S. and McLean, A.R. (2001) Infectious disease dynamics: what characterizes a successful invader? *Philosophical Tranactions of the Royal Society of London B* **356**, 901-910.

Newman, M.E.J. (2002) Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442.

Wong, C.W., Smith, S.E., Thong, Y.H., Opdebeeck, J.P. and Thornton, J.R. (1992) Effects of exercise stress on various immune functions in horses. *Am. J. vet. Res.* **53**, 1414-1417.

Zanette, D.H. and Kuperman, M. (2002) Effects of immunization in small-world epidemics. *Physica A* **309**, 445-452.