Network harness: bundles of routes in public transport networks

B. Berche*, C. von Ferber†,** and T. Holovatch1 *,†

*Statistical Physics Group, Institut Jean Lamour, UMR CNRS 7198, Nancy Université, 54506 Vandœuvre les Nancy Cedex, France
†Coventry University, Applied Mathematics Research Centre, CV1 5FB Coventry, UK
**Universität Freiburg, Physikalisches Institut, D-79104 Freiburg, Germany

Abstract. Public transport routes sharing the same grid of streets and tracks are often found to proceed in parallel along shorter or longer sequences of stations. Similar phenomena are observed in other networks built with space consuming links such as cables, vessels, pipes, neurons, etc. In the case of public transport networks (PTNs) this behavior may be easily worked out on the basis of sequences of stations serviced by each route. To quantify this behavior we use the recently introduced notion of network harness. It is described by the harness distribution \( P(r,s) \): the number of sequences of \( s \) consecutive stations that are serviced by \( r \) parallel routes. For certain PTNs that we have analyzed we observe that the harness distribution may be described by power laws. These power laws observed indicate a certain level of organization and planning which may be driven by the need to minimize the costs of infrastructure and secondly by the fact that points of interest tend to be clustered in certain locations of a city. This effect may be seen as a result of the strong interdependence of the evolutions of both the city and its PTN.

To further investigate the significance of the empirical results we have studied one- and two-dimensional models of randomly placed routes modeled by different types of walks. While in one dimension an analytic treatment was successful, the two dimensional case was studied by simulations showing that the empirical results for real PTNs deviate significantly from those expected for randomly placed routes.

Keywords: complex networks, harness effect, public transport
PACS: 02.50.-r, 07.05.Rm, 89.75.Hc

INTRODUCTION

A variety of different phenomena have in recent years been analyzed in the context of complex network theory [1, 2]. Usually, the focus is on the network topology while the study of specific features as e.g. the network load or real-space correlations is mostly left aside

Analyzing statistical properties of public transport networks (PTNs) [3, 4, 5] a so-called harness

1 holtaras@lpm.u-nancy.fr
2 Paper presented at the Conference “Statistical Physics: Modern Trends and Applications” (23-25 June 2009, Lviv, Ukraine) dedicated to the 100th anniversary of Mykola Bogolyubov (1909-1992).
effect which results from such spatial correlations has recently been proposed [5, 6]. The latter may be observed for networks on which a set of walks or paths is defined. Indeed, public transport routes sharing the same grid of streets and tracks define such a set of walks on this grid. Often these are found to proceed in parallel along shorter or longer sequences of stations. Similar phenomena are observed in other networks built with space consuming links such as cables [7], vessels [8], pipes [9], neurons [10], etc. In the case of PTNs, a quantitative description of sequences of stations that are served by several routes may be performed in the form of the harness distribution \( P(r,s) \): the number of maximal found on the network sequences of \( s \) consecutive stations that are serviced by \( r \) parallel routes. In a recent empirical analysis of the harness effect on PTNs of 14 major cities in the world [5, 6] it was found that for certain cities the behavior of the function \( P(r,s) \) is described by power laws. These observed power laws indicate a certain level of organization and planning which may be driven by the need to minimize the costs of infrastructure and secondly by the fact that points of interest tend to be clustered in certain locations of a city.

In the present paper, to further investigate the significance of the empirical results we have studied one- and two-dimensional models of randomly placed routes modeled by different types of walks. The setup of the paper is the following. First we recall some results of the empirical analysis. We then proceed solving a simple 1d model of a growing non-correlated network that allows an analytical treatment. Finally we analyze different types of growing networks on the 2d square lattice and perform numerical simulations and measure the resulting harness effect as function of the model parameters. Conclusions are given in the last section.

RESULTS OF AN EMPIRICAL ANALYSIS

In a previous study we have analyzed the PTNs of 14 major cities that have different geographical, historical and cultural background [5, 6]. For some of these we observed non-vanishing harness distributions \( P(r,s) > 0 \) even for long sequences \( s \) and high numbers of routes \( r \). This is what we call a "strong" harness effect (examples are Sao Paolo, Hong Kong, Istanbul, Los Angeles, Rome, Sydney, Taipei, Moscow, London; some are shown in Fig. 1). For other PTNs the maximal values of \( s \) and \( r \) with \( P(r,s) > 0 \) were found to be smaller than 10 (Berlin, Paris, Dallas, Duesseldorf, Hamburg) - this we call a "weak" harness effect (Fig. 2). It is important to note that the division into these two classes does not correlate with either the average number of routes \( R \) in the PTN or their average length \( S \). Another result is that similar to the node-degree distributions [5, 6] we observe that the harness distribution \( P(r,s) \) for some of the cities (Sao Paolo, Hong Kong, Istanbul, Los Angeles, Rome, Sydney) may be described by a power law:

\[
P(r,s) \sim r^{-\gamma_{s}}, \quad \text{for fixed } s,
\]

whereas the PTNs of other cities (Taipei, Moscow, London, Hamburg) are better described by an exponential decay:

\[
P(r,s) \sim e^{-r/\hat{r}_{s}}, \quad \text{for fixed } s.
\]

We illustrate this behaviour in Figs. 1a,b showing the harness distribution for Istanbul and for Taipei. In some cases (e.g. for Rome and Los Angeles) there is a crossover between regimes (1) and (2) at larger \( s \) as shown this for the PTN of Los Angeles (Fig.1c). Here, one can see that for small values of \( s \) the results are better described by a power law (1). With increasing \( s \) a tendency to an exponential decay (2) appears (Fig. 1d). This is less obvious for other cities analyzed, however in
all cases the harness distribution $P(r,s)$ as function of $r$ decays faster for longer sequence lengths and while also attaining a more pronounced curvature.

Note that in Fig. 1 we plot $s$-cumulative distributions $P_c(r,\hat{s})$ where a sequence with maximal length $s = 9$ will be counted once as a sequence of length $\hat{s} = 9$ and twice as a sequence of length $\hat{s} = 8$ etc:

$$P_c(r,\hat{s}) = \sum_{s=\hat{s}}^S (s + 1 - \hat{s})P(r,s)$$

It may be surprising that these curves e.g. for Taipei (Fig. 1b) intersect for low values of $r$. We will discuss this effect below.

For PTNs for which the harness distribution follows a power law (1) the corresponding exponents $\gamma_s$ are found in the range of $\gamma_s = 2 \div 4$. For those distributions with an exponential decay the scale $\hat{r}_s$ (see eq.(2)) varies in a range $\hat{r}_s = 1.5 \div 4$. The power laws observed for the behavior of $P(r,s)$ indicate a certain level of organization and planning which may be driven by the need to minimize the costs of infrastructure and secondly by the fact that points of interest tend to be clustered in certain locations of a city. Note that this effect may be seen as a result of the strong interdependence of the evolutions of both the city and its PTN. We want to emphasize that the harness effect is a feature of the network given in terms of its routes but it is invisible in any of the complex network representations of public transport networks presented so far, such as L-space [3], P-space [11] or
It is possible, that the notion of harness may be useful also for the description of other networks with similar properties. On the one hand, the harness distribution is closely related to distributions of flow and load on the network. On the other hand, in the situation of space-consuming links (such as tracks, cables, neurons, pipes, vessels) the information about the harness behavior may be important with respect to the spatial optimization of networks. A generalization may be readily formulated to account for real-world networks in which links (such as cables) are organized in parallel over a certain spatial distance. While for the PTN this distance is simply measured by the length of a sequence of stations, a more general measure would be the length of the contour along which these links proceed in parallel.

For the cities observed no correlation appears to occur between the harness distribution behavior and other well-known network characteristics that were analyzed, as for example the node-degree distribution of PTNs [5].

FIGURE 2. $s$-cumulative harness distribution $P_c(r, s)$ as function of $r$ for fixed $s$ for Paris ($s = 2 \div 8$). $P_c(r, s) = 0$ if $s > 2$ for $r > 7$ and if $s > 7$ even for $r > 2$.

However, the extent to which harness properties are expressed may obviously play a role for the attack vulnerability of a PTN. Interestingly, our previous investigations have shown that the Paris PTN is most resilient to any type of random or directed attacks (in terms of percolation concepts) among all analyzed PTNs [13]. At the same time it exhibits the "weakest" behavior with respect to the harness effect (Fig. 2). One may expect such a result: routes that do not share the same streets are more resilient. However, for other cities no apparent correlation between harness effect behavior and their vulnerability has been found so far.

As noted above the interesting question to answer is: are there any structural evolution purposes behind this effect, or can it be found just as well within simple random scenarios.

**ANALYTIC RESULTS AND MODELING IN 1D**

Let us first investigate a network model with routes placed randomly in one dimensional space. Although being very simple this model can mimic a harness effect and as we will see below, it allows an analytical solution. The model is formulated in the following way:

The left terminals of $R$ routes of length $S$ are placed at random on a line with $N$ sites, with periodic boundary conditions. E.g. in Fig.3 we show two routes of length $S = 15$ with left terminals at $x = 0$ and $x = -8$. We define the route density as $\rho = R/N$ routes per site.
FIGURE 3. $R = 2$ routes given as simple sequences of $S = 15$ consecutive sites are placed at random on a line with $N$ sites.

The distribution of left-terminals on a given site, e.g. site $x$ will be $P_x(r)$: the probability that $r$ routes have their left-terminal on site $x$

$$P_x(r) = \binom{R}{r} \cdot \left(\frac{1}{N}\right)^r \cdot (1 - \frac{1}{N})^{R-r},$$

where the first term counts the number of ways to select $r$ of $R$ routes, the second term is the probability that the $r$ left terminals lie on site $x$ and the third term is the probability that no left terminal of an unselected route lies on site $x$. In other words, by definition $P_x(r)$ is a binomial distribution. For $N \to \infty$, but fixed $\rho = R/N$ this distribution has the limiting behavior of a Poisson distribution:

$$P_x(r) \approx e^{-\rho} \cdot \frac{\rho^r}{r!}. \quad (5)$$

Let us now calculate the probability that there is a sequence of maximal length $s$ and maximal width $r$ of $r$ routes in parallel between sites $x = 0$ and $x = s$. This implies that at least one of the $r$ routes starts at $x = 0$ and at least one of the routes ends at $x = s - 1$. The latter route then starts at $x = -(S - s) \equiv -\delta$ (Fig. 3). The other $r - 2$ routes may start anywhere in between $-\delta \leq x \leq 0$. We denote the number of routes starting at $x = 0$ as $r_0$, those starting at $x < 0$ as $r_{-x}$.

In the limit $R \gg r$ and $N \gg \delta$ we may consider $P_0(r_0), P_1(r_1), \ldots, P_s(r_s)$ as independent probabilities (this is not true for small systems, however, as we see later the correlation between these probabilities is negligible for the cases studied here. The overall probability to find a sequence of length $s$ and width $r$ starting at $x = 0$ is then the sum over all combinations leading to the result:

$$P_0(r,s) = \sum_{\{r_i\}, r_0 \geq 1, r_s \geq 1}^{(r)} P(r_0) \cdot P(r_1) \cdot \ldots \cdot P(r_s) = \sum_{\{r_i\}, r_0 \geq 1, r_s \geq 1}^{(r)} e^{-\delta p} \cdot \frac{\rho^{r_0}}{r_0!} \cdot \frac{\rho^{r_1}}{r_1!} \cdot \ldots \cdot \frac{\rho^{r_s}}{r_s!} = e^{-\rho \delta} \cdot \rho^r \cdot \sum_{\{r_i\}, r_0 \geq 1, r_s \geq 1}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!}. \quad (6)$$

where $\sum_{\{r_i\}}^{(r)}$ denotes a sum over $\{r_i\}$ with $r = r_0 + r_1 + \ldots + r_s$.

Now, without the conditions $r_0 \geq 1$ and $r_s \geq 1$ this sum can be derived from:

$$(\delta + 1)^r = (1 + 1 + 1 + \ldots + 1)^r = \sum_{\{r_i\}}^{(r)} \frac{r!}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!}. \quad (7)$$

The sum with these conditions however can be written as:

$$\sum_{\{r_i\}, r_0 \geq 1, r_s \geq 1}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!} = \sum_{\{r_i\}}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!} - \sum_{\{r_i\}, r_0 = 0, r_s \geq 1}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!} - \sum_{\{r_i\}, r_0 \geq 1, r_s = 0}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_s!} = \frac{(\delta + 1)^r}{r!} - 2 \frac{(\delta)^r}{r!} + \frac{(\delta)^r}{r!}. \quad (8)$$
Thus:

\[ P_0(r,s) = e^{\left( -\bar{s} \rho \right)} \cdot \rho^r \cdot \left[ (\bar{s} + 1)^r - 2(\bar{s})^r + (\bar{s} - 1)^r \right] / r! . \]  

(9)

With this formula we count sequences that start at \( x = 0 \). To receive the overall probability this is to be multiplied by \( N \).

\[ P(r,s) = N \cdot e^{\left( -\bar{s} \rho \right)} \cdot \rho^r \cdot \left[ (\bar{s} + 1)^r - 2(\bar{s})^r + (\bar{s} - 1)^r \right] / r! . \]  

(10)

Simple arithmetic allows us to calculate the \( s \)-cumulative distributions \( P_c(r,\bar{s}) \), see eq. (3).

As mentioned above, the probabilities we use are appropriate for infinite systems. To test their validity for finite cases we performed some simple simulations. It turns out that the average results fit this formula with very good accuracy even for small \( N \), for example \( N = 10 \) and of course for any larger values of \( N \) (Fig. 4a). For a large range of parameters \( R, S, N \) the behavior of \( P(r,s) \) looks similar to what is shown in Fig. 4b. For high overall density \( \rho \cdot S \) we observe that curves for different \( s \) intersect at small \( r \) as for some real-world PTNs. In one dimension, this effect has the following explanation: when the space is overcrowded one will in general find more than two routes to overlap for small sequences of stations.
This model has three parameters: the number of routes $R$, the route length $S$ and the number of sites $N$. As is obvious from eq. (10) the harness distribution $P(r,s)$ for all $r$ and $s$ is close to a Poisson decay (5) for any set of parameters. Therefore PTNs with an exponential behavior of the harness distribution $P(r,s)$ may be compared with the results of the one-dimensional approach. As an example we compare the normalized harness distribution for Moscow and the one-dimensional set of lines (Fig. 4c,d), where the number of routes $R$ and the route length $S$ were chosen to match those of the Moscow PTN ($S$ set to the average route length).

However, the quantitative results for all observed PTNs are several orders of magnitude higher than the result obtained with (9) in one dimension for the same $R$ and $S$ (for any $N$). Furthermore, PTNs that show a power law behavior (1) are even qualitatively different from the random 1D approach. Another difference is that the harness distribution curves for different $s$ are very similar in shape and both slope and curvature vary much less than for the PTN harness distributions.

In the following we test a two-dimensional model with the simple simulations.

**SIMPLE 2D MODELING**

It is obvious that simply throwing at random lines parallel to the axis’ of a 2d square lattice with periodic boundary conditions will lead to the original 1d problem: If the lattice has $X \times X$ sites one would get $2X$ independent one-dimensional systems. However, it is not a priory clear what results one will find for more general sets of walks on a 2d square lattice. To work this out,
we implemented the following simulations. We work on a 2d $X \times X$ square lattice with periodic boundary conditions. On this lattice we chose a set of $R$ walks each of length $S$ (number of steps plus 1). The routes are built either as random walks (RW), non-reversal random walks (NRRW) that cannot reverse the previous step, or self-avoiding random walks (SAW), that may not intersect themselves.

These models have three parameters: the number of routes $R$, the route length $S$ and the lattice size $X$. We choose the first two parameters to match those of different real PTNs.

Postponing a more detailed analysis to a separate publication we here summarize some of the main features of the harness distribution $P_{cc}(r, \hat{s})$ of these models. Besides the finding that the harness effect is "weak", some similarity between the the harness effects seen in the three models is observed (Fig 5a,b,c). Curvature and slope evolve in a similar way. Also intersections between the curves for different $s$ are found to occur at lower values of $r$ in all cases. Differences are that the RW-generated networks demonstrate a "weaker" harness effect, while NRRW- and SAW-generated networks result in harness distributions $P_{cc}(r, \hat{s})$ of similar order of magnitude.

It turns out that for fixed $R$ and $S$, increasing the lattice size $X$, the harness distributions $P(r, s)$, for all fixed $r < R$ and $s < S$ show non-monotonous behavior (Fig. 6). As function of $X$ it first increases, and then after reaching a maximum it starts to decrease. Comparing with the empirical values found for real PTNs we observe, that for some PTNs the empirical values even for small $r$ and $s$ are significantly larger than the maximum that could be obtained with RW. For NRRW or SAW the empirical values are within the observed range, however only for a small interval of $X$.

This proves numerically the not surprising observation, that with any of the proposed random or quasi-random walks only a very "weak" harness effect may be obtained. In turn, this strongly indicates that for most of the observed cities the harness effect must have a structural background, that is not to be modeled by any of the random approaches taken here.

Let us therefore analyze the harness effect for a model that intrinsically takes correlations into account using some ideas of non-equilibrium network evolution. [5].

**AN EVOLUTIONARY MODEL IN 2D**

From previous studies of PTNs [4, 5] the following facts are relevant for our further discussion of PTN network evolution. The final network often displays features (a power law degree distribution)
FIGURE 7. Simulated PTN maps. \( X = 300, R = 1024, S = 64 \). Parameter \( a = 0 \). a) \( b = 0.1 \) b) \( b = 0.3 \) a) \( b = 0.5 \). One can see how parameter \( b \) influences the spread of the network.

that are compatible with a preferential attachment scenario [1]. Furthermore, the fractal dimension of public transport routes appears to match that of 2d self-avoiding walks [5]. This leads to the following model.

Given a 2d square lattice with side \( X \) and periodic boundary conditions, \( R \) routes each of length \( S \) will be created in the following way. A first route is built as self-avoiding walk. Let \( k_x \) be the number of visits to lattice site \( x \). The \( R - 1 \) subsequent routes are constructed as SAWs with the following preferential attachment rules:

a) choose a terminal station at \( x_0 \) with probability

\[
p \sim k_{x_0} + a/X^2; \tag{11}\]

b) choose any subsequent station \( x \) of the route with probability

\[
p \sim k_x + b. \tag{12}\]

c) If route intersects itself, discard it and return to step a).

Repeat steps a)-c) until \( R \) routes are created.

We call this model the model of mutually interacting self-avoiding walks (MI SAW). It implements a preferential attachment scenario for networks that are built from chains (routes in our case). On the figures presented (Fig. 7-8) one can see how the parameters \( a \) and \( b \) influence the distribution of the routes on the lattice. Parameter \( a \) controls the number of disconnected clusters, while parameter \( b \) is related to the spread of each cluster. If both parameters equal zero, all \( R \) routes are restricted to the sites occupied by the first route.

To present the obtained results we show the harness distribution \( P_c(r, \hat{s}) \) for different values for the model parameters \( a \) and \( b \) (Fig. 9). It turns out that for choosing different values we may
either find a behavior that corresponds to a power law distribution (Fig. 9a,b) or a behavior that corresponds to an exponential decay (Fig. 9c,d).

Also other features of real-world PTNs may be reproduced. With increasing the sequence length \( s \) the harness distribution \( P(r, s) \) as a function of the number of routes \( r \) attains a stronger curvature and a steeper slope. Thus, with increasing sequence length \( s \) there is a cross-over from the power law regime (1) to an exponential regime (2).

In the same way as observed for real-world PTNs in the 1d approach and also for simple 2d models on the square lattice the \( s \)-cumulative harness distributions \( P_c(r, \hat{s}) \) as function of the number of routes \( r \) intersect at low values of \( r \). In one dimension we observed this behaviour because of an overcrowding effect, see above. This also occurs to some extent in the 2d case, however, on the 2d lattice there is an additional combinatorial effect that leads to this result: the number of different possible configurations of sequences with length 2 is smaller than the corresponding number of different possible configurations with length 3. To summarize one can say that the MI SAW model reproduces a large amount of the empirically observed behavior of harness distributions of PTNs.
CONCLUSIONS

Harness phenomena may be observed in different networks built with space consuming links such as cables, vessels, pipes, neurons, etc. The present analysis may possibly be applied also to such types of networks. In the particular case of the PTNs that we have analyzed we observe that in some cases the harness distribution may be described by power laws. These observed power laws indicate a certain level of organization and planning which may be driven by the need to minimize the costs of infrastructure and secondly by the fact that points of interest tend to be clustered in certain locations of a city. This effect may be seen as a result of the strong interdependence of the evolutions of both the city and its PTN.

To further investigate the significance of the empirical results we have studied one- and two-dimensional models of randomly placed routes modeled by different types of walks. While in one dimension an analytic treatment was successful, the two dimensional case was studied by extensive simulations.

Our main results are the following:

• A one dimensional model for harness distributions was solved analytically.
• Exponentially decaying harness distributions may be reproduced by the 1d approach.
• Simple random placement of RW, SAW or NRRW on a two dimensional square lattice result in weak harness distributions; in the RW case much weaker than for real PTNs.
• The $s$-cummulative distributions for different $s$ intersect at low values of $r$ for all models due to combinatorial reasons.
• A model of mutually interacting SAWs reproduces many of the empirically observed features of harness distributions.

ACKNOWLEDGMENTS

We thank Yu. Holovatch for comments on the manuscript and on the problem in general. T.H. is fully supported by Ecole Doctorale EMMA.

REFERENCES

1. R. Albert, and A.-L. Barabási, *Statistical Mechanics of Complex Networks*, Rev. Mod. Phys., **74**, pp. 47-97, (2002).

2. S. N. Dorogovtsev, and J. F. F. Mendes. *Evolution of Networks*, Oxford University Press, Oxford, (2003).

3. V. Latora, and M. Marchiori, *Is the Boston subway a small-world network*, Physica A, **314**, pp. 109-113 (2002)

4. J. Sienkiewicz, and J. A. Holyst, *Statistical analysis of 22 public transport networks in Poland*, Phys. Rev. E, **72**, 046127, (2005).

5. C. von Ferber, T. Holovatch, Yu. Holovatch, and V. Palchykov, *Network harness: Metropolis public transport*, Physica A, **380**, pp. 585-591 (2007); *Public transport networks: empirical analysis and*
6. C. von Ferber, Yu. Holovatch, and V. Palchykov, *Scaling in Public Transport Networks*, Condens. Matter Phys., 8, No. 1(41), pp. 225-234, (2005).

7. M. R. Cutkosky, A. B. Conru, and S-H. Lee, *An agent-based approach to concurrent cable harness design*, AIEDAM, 8, No. 1, (1994).

8. P. Carmeliet, and M. Tessier-Lavigne, *Common mechanisms of nerve and blood vessel wiring*, Nature, 436, pp. 193-200, (2005).

9. N. Hwang, and R. Houghtalen, *Fundamentals of hydraulic Engineering Systems*, Prentice Hall, Upper Saddle River, NJ, (1996).

10. J. G. White, E. Southgate, J. N. Thompson, and S. Brenner, *The structure of the nervous system of the nematode C. Elegans*, Philos. Trans. Roy. Soc. London, 314, pp. 1–340, (1986).

11. P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, *Small-world properties of the Indian railway network*, Phys. Rev. E, 67, 036106, (2003).

12. J.-L. Guillaume, and M. Latapy, *Bipartite graphs as models of complex networks*, Physica A, 317, pp. 795-813 (2006);

13. B. Berche, C. von Ferber, T. Holovatch, and Yu. Holovatch, *Resilience of public transport networks against attacks*, arXiv:0905.1638v1; Eur. Phys. J. B, submitted, (2009).