PROPERTY GRAPHS – A STATISTICAL MODEL FOR FIRE AND EXPLOSION LOSSES BASED ON GRAPH THEORY

BY

PIETRO PARODI AND PETER WATSON

ABSTRACT

It is rare that the severity loss distribution for a specific line of business can be derived from first principles. One such example is the use of generalised Pareto distribution for losses above a large threshold (or more accurately: asymptotically), which is dictated by extreme value theory. Most popular distributions, such as the lognormal distribution or the Maxwell-Boltzmann-Bose-Einstein-Fermi-Dirac (MBBEFD), are convenient heuristics with no underlying theory to back them. This paper presents a way to derive a severity distribution for property losses based on modelling a property as a weighted graph, that is, a collection of nodes and weighted arcs connecting these nodes. Each node v (to which a value can also be assigned) corresponds to a room or a unit of the property where a fire can occur, while an arc $(v, v'; p)$ between vertices v and v' signals that the probability of the fire propagating from v to v' is p. The paper presents two simple models for fire propagation (the random graph approach and the random time approach) and a model for explosion risk that allow one to calculate the loss distribution for a given property from first principles. The MBBEFD model is shown to be a good approximation for the simulated distribution of losses based on property graphs for both the random graph and the random time approach.

KEYWORDS

Random graph, exposure curve, MBBEFD, critical threshold.

1. INTRODUCTION

In an ideal world, loss curves should be derived from first principles, or there should be significant empirical evidence for their efficacy. This is how successful models are developed in science. As an example, consider Planck's distribution of density $f(x) = cx^3 \frac{1}{\exp(ax) - 1}$. It was introduced by Planck to describe the

Astin Bulletin 49(2), 263-297. doi:10.1017/asb.2019.4 © 2019 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

spectral distribution of blackbody radiation and was successfully explained by Einstein from first principles by assuming that radiation is made of quantised harmonic oscillators (Brehm and Mullin, 1989).

Other examples include the normal (or Gaussian) distribution $f(x) = cx^{-(x-\mu)^2/2\sigma^2}$, which can be proven (central limit theorem) to be the limiting distribution for the empirical mean (Freund, 1999), and the Pareto distribution, $f(x) = cx^{-(\alpha+1)}$, $x > 1$, which can be proven to describe the distribution of the size of clusters in fractals (Mandelbrot, 1982).

Even when such clean derivations are not available, a specific distribution can be empirically proven to be effective in a given domain. For example, the Pareto distribution is effective in modelling the distribution of income (Pareto, 1964) and was actually introduced in that context.

As opposed to this, severity distributions in actuarial practice tend to be chosen for heuristic or historical reasons and on the basis of distribution fitting starting from a dictionary of typical distributions such as Lognormal, Gamma or Burr. This causes over-fitting and produces models of limited predictive value (Chhabra and Parodi, 2010; Parodi, 2014, Section 16.2). When enough losses are available, these distributions appear often inadequate to represent the underlying loss generating process.

One exception to this is the use of the generalised Pareto distribution for modelling the tail of severity distributions, which is solidly grounded on results from extreme value theory (Embrechts *et al.*, 1997).

In this paper we explore situations where loss distributions can be derived from first principles for property risk, based on a simplified model of a property structure based on graph theory. Graph theory is the mathematical underpinning for the study of networks, that is, a configuration of points (nodes or vertices) with pairwise connections (arcs or edges), possibly weighted (Frieze, 2015). The underlying idea is that properties can be broken down into units that are connected by an arc if there is a way for fire to propagate from one unit to another; the connection is more or less strong depending on the ease by which the loss propagates. The loss is then modelled as the sum of the values of those units.

Armed with such a model, we can then try to derive a statistical distribution of the size of the fire losses, and from that the corresponding exposure curve. An exposure curve (see Appendix A) is a curve in 1:1 correspondence with the severity curve that allows us to assess the impact of different retention levels and to allocate losses to the layers (Bernegger, 1997; Riegel, 2010; Parodi, 2014, Chapter 21). We can also try to predict how the shape of the exposure curve is affected by the topological structure of the property when we are presented with new types of property. Finally, and especially for large commercial properties, we can decide (if we want) to use a curve that reflects the idiosyncrasies of a specific property rather than adopting the default portfolio curve.

Needless to say, this study is only an initial foray into the graph-theoretic approach, and further work is needed to fully develop it.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

1.1. Overview of related literature

The authors are not aware of any other attempts to apply the theory of random graphs and phase transitions to the production of exposure curves in property insurance. However, the possibility of using random graphs in connection with fire propagation has been noticed before, especially in relation with forest fire (Rath and Toth, 2009).

There is a wide body of literature on models of fire propagation and explosion risk for specific circumstances and structures from a scientific and engineering viewpoint. We cannot attempt here to give an exhaustive or even sufficient list of references. However, a good starting point for fire is perhaps the material on fire dynamics published by the US National Institute of Standards and Technology (NIST) on their website[1]. The NIST has also produced a fire dynamics simulator. The question of the probability of fire propagation between property units can also be related to Layer of Protection Analysis (LOPA) (Center for Chemical Process Safety, 2010; Willey, 2014) and the Swiss Cheese model (Reason, 1990).

There is also a vast literature on explosion modelling. A good place to start is the documentation related to explosion modelling tools to which we refer in this paper, such as Swiss Re's ExTool[2]. Examples of more complex models that require more detailed data inputs are PHAST, Baker–Strehlow–Tang (BST) by BakerRisk, Multi Energy Model by TNO, the Netherlands Organisation for applied scientific research.

The methodology for the derivation of loss models as presented here is based on the theory of random graphs, which was introduced and developed by Erdös and Rényi and is presented at length in Bollobas (2011), Frieze (2015) and in summary format in Slade (2008).

The basics of graph theory, which are used to calculate the relevant properties of the property graphs (e.g., size of the maximum connected cluster), are discussed, for example, in Trudeau (2003).

Exposure rating for property business and the Maxwell-Boltzmann-Bose-Einstein-Fermi-Dirac (MBBEFD) parameterisation (originally introduced by Bernegger (1997)) is discussed, for example, in Riegel (2010) and Parodi (2014, Chapter 21).

The results presented here make some use of UK Government material on building regulations (UK Government, 2010) and statistics from the Office for National Statistics (National Archives, 2011).

2. The random graph approach to fire propagation

We are going to look at two alternative ways of approaching fire propagation. Common to these two ways is the idea of representing the property as a weighted graph. Even before defining what a graph is, it is useful to have an intuitive idea of what this representation might look like.

Figure 1 shows an example of a typical three-bedroom house in the UK. Each room (or, more generally, "property unit") can be represented as a node

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

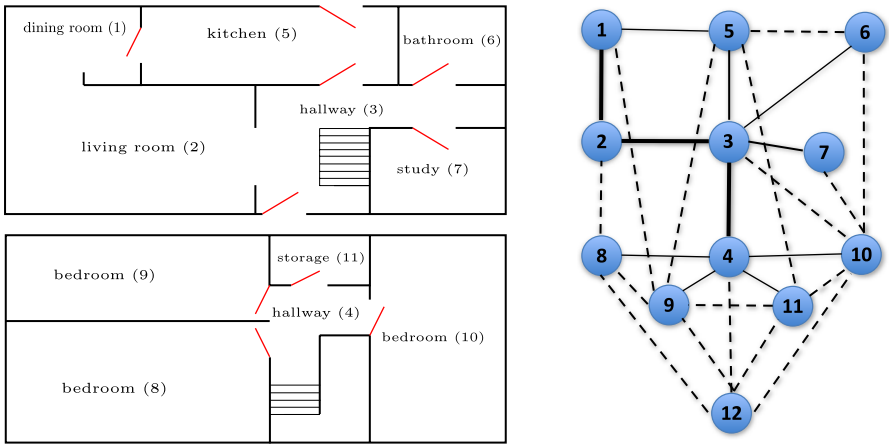


FIGURE 1: Graph representation of a property (example). (Left) An example of a typical three-bedroom house in the UK. (Right) The corresponding graph representation (node 12 is a loft, not shown in the left panel). Also not shown is that between the living room (node 2) and the kitchen (node 5), there is a fireplace, so that direct propagation via this route is prohibited and there is therefore no arc connecting these two nodes.

in a network (graph) and two nodes are connected if there is a pathway between the two corresponding rooms through which fire can propagate. In this particular case, the thickness of a line represents the ease through which fires can propagate: a thick continuous line when two units are not segregated; a thin continuous line when two units are separated by a door; a thin dashed line when two units are separated by a wall or floor/ceiling.

To see how this representation can help us, let us start with a quick excursus of those parts of graph theory that are relevant to us.

2.1. Graph theory – a quick primer

A graph is a powerful mathematical representation of systems that are made of interconnected parts. Since complex systems are ubiquitous, graph theory (a branch of discrete mathematics) finds applications to a large number of disciplines, from engineering to telecommunications to software engineering. Your sat nav uses (among other things!) a graph representation of a road atlas that allows it to solve efficiently problems such as "what is the fastest route between A and B?".

For the purpose of this paper we only need a few basic concepts of graph theory. The reader interested in knowing more about the subject is referred to Trudeau (2003) for the basic concepts of graph theory and Bollobas (2011) and Frieze (2015) for an introduction to random graph theory. Slade (2008) provides a succinct description of both topics.

2.1.1. *Basic definitions.*

Here is a list of definitions which will be necessary to proceed.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

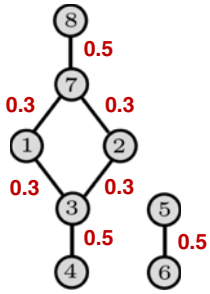


FIGURE 2: An example of a weighted graph with two connected components.

A *graph* is an object consisting of two sets: a set of *nodes* and a set of *arcs* connecting pairs of nodes. For example, nodes might represent cities and arcs might be roads connecting these cities.

A *complete graph* is a graph in which there is an arc between all pairs of nodes. This means that if there are n nodes, a complete graph will have $n(n-1)/2$ arcs.

A *weighted graph* is a graph where every arc has an associated weight. In the example above with cities and roads, the weight may be the length of the road connecting two cities.

The most important property of graphs for the purpose of this paper is that of connectedness.

Two nodes v, v' are said to be *connected* if there is a *path* from v to v', that is, a sequence of adjacent arcs from v to v'.

A *connected component of a graph* is a subgraph such that any two nodes in the subgraph are connected.

A directed graph, or *di-graph*, is a graph in which the arcs have a direction, that is, it matters whether the arc goes from v to v' or from v' to v. The arcs are in this case represented by arrows. Di-graphs can also be weighted.

Obviously, these basic ideas can be extended in a number of directions: a useful extension for us is that nodes (and not only arcs) can be given a value or a label.

Figure 2 shows an example of a weighted graph with two connected components.

2.1.2. *Classical problems.*
Classical graph-theoretic problems are

- find the shortest path between two nodes;
- determine if a graph is connected;
- determine whether the graph contains a cycle (i.e. a path that starts and ends at the same node);
- find the shortest cycle that visits all nodes (travelling salesman problem); and
- can a graph be drawn on a plane without the arcs intersecting except at nodes (i.e., is a graph planar)?

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

```
1 → (3, 0.3) → (7,0.3)
2 → (3, 0.3) → (7, 0.3)
3 → (1, 0.3) → (2, 0.3) → (4, 0.5)
4 → (3, 0.5)
5 → (6, 0.5)
6 → (5, 0.5)
7 → (1, 0.3) → (2, 0.3) → (8, 0.5)
8   (7, 0.5)
```

FIGURE 3: Adjacency list representation of a weighted graph.

2.1.3. *Graph representations.*
For computational purposes, a graph can be represented in different ways. Two popular ones are as adjacency lists and as adjacency matrices.

Adjacency list. This lists all the nodes in the graph and the nodes connected to each of the nodes, along with the weight of the arc. For example, the graph in Figure 2 can be represented as the adjacency list in Figure 3. Without entering into computer-science technicalities, note that the symbol "→" should be interpreted as something akin to "points to" and simply specifies how two items are stored in memory. For example, $(3, 0.3) \rightarrow (7, 0.3)$ does not mean that 3 and 7 are connected by an arc but that their memory locations are cross-referenced. In the case of an unweighted graph this representation becomes even simpler, for example, the first row would look like $1 \rightarrow 3 \rightarrow 7$.

For a di-graph the representation is exactly the same, except that it may be that a node be in the list of another node but not vice versa, and that the weight in the two directions may be different.

This representation is memory-effective if the number of arcs in the graph is small compared to the number of arcs in a complete graph, which equals $n(n-1)/2$, where n is the number of nodes, double that for a di-graph.

Adjacency matrix. Another graph representation which is quite intuitive is that of a matrix. Here, the nodes appear both in the rows and in the columns of the matrix and the matrix element at the intersection is >0 whenever there is an arc connecting the nodes. For example, for the graph in Figure 2 the representation is as in Figure 4.

Note that for an undirected weighted graph, the adjacency matrix is symmetrical as that of Figure 4. For a di-graph, the adjacency matrix is in general non-symmetrical. A number different from zero in the diagonal would indicate an arc from a node to itself.

2.1.4. *Random graphs.*
We are especially interested in the concept of "*random graphs*". These are graphs that are built by connecting at random a given number of nodes. There are two processes for building random graphs: you can start from n nodes none of which are connected, and add arcs between nodes at random; or you can start from a complete graph and remove arcs at random. For the purpose of

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

	1	2	3	4	5	6	7	8
1	0	0	0.3	0	0	0	0.3	0
2	0	0	0.3	0	0	0	0.3	0
3	0.3	0.3	0	0.3	0	0	0	0
4	0	0	0.3	0	0	0	0	0
5	0	0	0	0	0	0.5	0	0
6	0	0	0	0	0.5	0	0	0
7	0.3	0.3	0	0	0	0	0	0.5
8	0	0	0	0	0	0	0.5	0

FIGURE 4: Adjacency matrix representation of an undirected weighted graph.

our paper we are interested in the second way of producing graphs, and especially in a modified version of it in which rather than starting from a complete graph we start from *any* graph and remove arcs at random.

2.2. The random graph approach – the basic idea

We are now in a better position to illustrate the approach that we briefly described at the beginning of Section 2. The idea is to represent the property as a weighted graph G, very much like that in Figure 1. Every node v in the graph represents a property unit, every arc a a possible passageway for fire, and the weight w_a of the arc (which was visualised in Figure 1 by arcs of different width/style) the probability that the fire actually spreads through that gateway.

The spread of a fire can then be simulated as follows. Please refer to Appendix A for the explanation of terminology related to exposure rating (e.g., damage ratio, normalised severity curve).

Calculation of normalised severity curve – pseudocode

```
Input: a graph G=(V,A,W) where V is the set of vertices, A is the set of arcs and W is the set of the weights associated with each arc.

Calculate maximum possible loss (MPL) as the size of the largest connected component of G.
For each simulated scenario j=1, ... N
    Pick a node v∈V at random - this is where the fire starts
    For each arc a∈A
        Delete a with probability =1−w_a, where w_a∈W is the weight associated with a
    Output for scenario j= a new (unweighted) graph G_j where
```

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

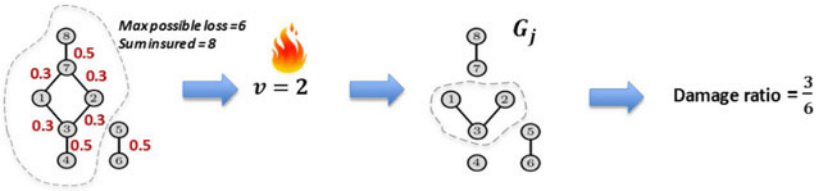


FIGURE 5: Illustration of the simulation process. Note that the MPL is equal to 6 while the whole insured property has 8 units. First the origin of the fire (node 2) is selected, then arcs are removed with probability equal to 1 minus the weight of that arc, then the size of the connected component that includes node 2 is calculated, and finally the loss is divided by the MPL, giving the damage ratio.

The nodes are the same as those of G.

The arcs are the surviving arcs, unweighted.

Loss for scenario j = Size of the connected component of G_j that includes v.

Damage ratio for scenario j: Damage_Ratio = Loss/MPL.

Sort damage ratios in ascending order to produce a normalised severity curve.

The simulation process can be visualised as in Figure 5. Note that the crucial (and hardest) part of the algorithm above is determining the connected components of the graph. This is done by well-known graph exploration techniques such as depth-first search and breadth-first search that are part of any graph theory package and specifically of the Python package NetworkX (Hagberg *et al.*, 2018) that we used for this paper.

2.3. Erdős–Rényi theory – a useful asymptotic result

The literature around random graphs yields some off-the-shelf results that are relevant to the simulation process outlined in Section 2.2. An important one (Slade, 2008) concerns the mean size of the connected component that includes a vertex v. The behaviour of the mean size depends on where the probability p of two nodes being connected stands with respect to a *critical threshold* $p_c = \frac{1}{n-1}$, where n is the number of nodes:

$$\mathbb{E}(\text{size}\,(p)\,) \sim \begin{cases} \frac{1}{|\varepsilon|} & p < \frac{1}{n-1} \\ \gamma n^{\frac{1}{3}} & p = \frac{1}{n-1}, \\ 4\varepsilon^2 n & p > \frac{1}{n-1} \end{cases} \tag{2.1}$$

where γ is a constant and $\varepsilon = p\,(n-1) - 1$. This is an asymptotic result for large n.

Note that when $p > p_c$ the expected size of the loss is of the same order of n, which in turn is of the same order of the MPL: $\mathbb{E}(\text{size}\,(p)\,) \sim \text{MPL}$. This is

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

the case that exposure curves (see Appendix A) are designed to capture, as it describes losses that are a fraction of the MPL. The case $p < p_c$ describes fires that are contained (propagation is unlikely), whose sizes are independent of the overall size of the graph, and can therefore be described as "attritional".

Although Equation (2.1) is interesting, the question we actually want answered is: what is the distribution of size(p)? Also, while there may be asymptotic results for the actual size distribution, we are interested in the case where the number of units is finite and possibly very small. We are not aware of any analytical results available for this case. In order to respond to this question, we have therefore resorted to Monte Carlo simulation. A discussion of the convergence property of our Monte Carlo simulation process that is valid for all simulations used in this paper can be found in Appendix D.

2.4. Simulating fires with random graphs

We are going to run a number of experiments to estimate the exposure curves from random graphs. We are first going to consider fires in *complete* graphs: this is an unrealistic model of properties but it helps us explore how the exposure curves change when we change the value of the probability of fire propagation (Section 2.4.1). After an aside on how we should judge the goodness of fit between the MBBEFD model and the empirical[3] exposure curves that emerge from random graph simulations (Section 2.4.2), we move to consider property graphs that actually reflect typical residential structures (Section 2.4.3).

2.4.1. *Simulating fires with complete graphs.*
The first experiment that we do is to start from a complete graph with $n = 20$ vertices and remove each of the edges with probability $1 - p$. We try different levels of probability: $p = 0.025$, $p = 0.05$ and $p = 0.1$.

We repeat this a large number of times and each time we select a vertex v at random and calculate the size of the connected component that contains v. We then divide the size of the connected component by the MPL, which in this case is equal to $n = 20$.

Based on the output of this simulation we can produce an empirical severity distribution and the corresponding exposure curve. We also attempt to fit these empirical curves with MBBEFD models (see Appendix A for their definition).

To do that, we have used a slightly different parameterisation than the usual (b, g) one as introduced by Bernegger. We have parameterised the MBBEFD curve in terms of parameters (k, l) that are related to (b, g) by these equations:

$$b = e^k, g = 1 + e^l. \tag{2.2}$$

The advantage of this new parameterisation is that while b and g may vary over different orders of magnitude, most exposure curves have k between -15 and 5 and l between -5 and 20. This makes it much easier to search

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

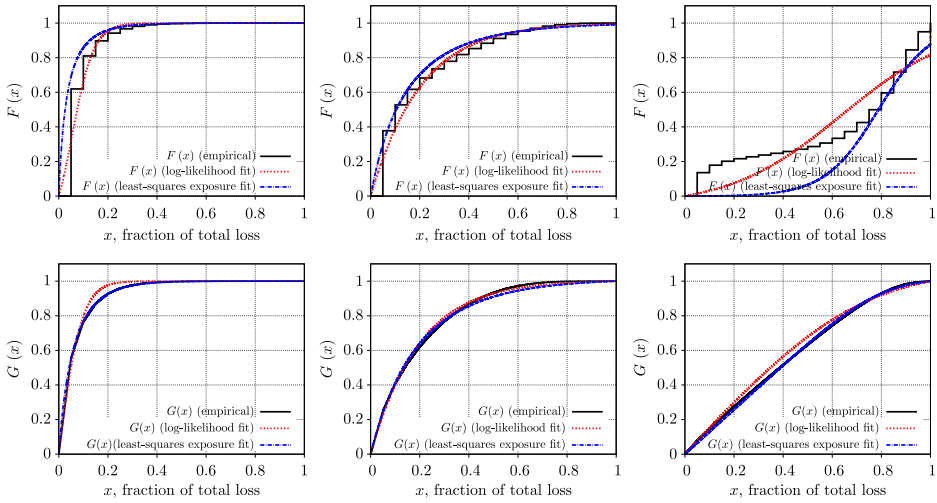


FIGURE 6: (Top) Empirical severity curve and MBBEFD fit using both MLE and least-squares regression (of the corresponding exposure curve) for different values of the propagation probability, starting from a complete graph. From left to right, p = 0.025 (< critical), p = 0.05 (≈ critical), and p = 0.1 (> critical). (Bottom) The corresponding MBBEFD exposure curves.

the parameter space with numerical optimisation algorithms such as the Sequential Least Squares Programming algorithm (SLSQP), part of the standard Python routine scipy.optimise.minimise that we have used here (Kraft, 1988). Throughout this paper, we have used two different ways of calibrating the parameters: maximising the log-likelihood of the cumulative distribution function ("log-likelihood fit") and doing a least-squares regression fit on the exposure curve directly ("least-squares exposure fit").

The results are shown in Figure 6, which shows examples in which the propagation probability is less than, approximately equal to or larger than the critical threshold $p_C = \frac{1}{19}$. Note that the most relevant case for our purposes is that in which the probability is larger than the critical value, because that is the case where losses will be of the same scale as the graph size. When the probability is lower than the critical threshold, the losses will tend to remain confined.

In all cases, the MBBEFD curves provide what appears to be a decent fit for the exposure curve (bottom) and (to a much lesser extent) for the severity curve (top). However, what constitutes a good fit for our purpose? We need to delve into this a bit deeper.

2.4.2. *Is the MBBEFD distribution a good fit for exposure curves based on random graphs?*

How do we judge the goodness of fit of the MBBEFD model to the empirical severity/exposure curves derived empirically (i.e. by stochastic simulation) from a given property graph?

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

One option is of course using one of the classical statistical tests, such as the Kolmogorov–Smirnov (KS) test for the hypothesis that two cumulative distribution functions are the same. However, a little thought reveals that this is bound to fail as we are comparing a continuous distribution (the MBBEFD severity curve) with one that has quite large jumps. They are two different distributions, and any sensible test will simply reveal that.

Of course we can use the KS distance anyway and simply demand that it is below a certain threshold, not necessarily related to the critical thresholds of the KS test. However, a more business-relevant way of comparing the model with our graph-derived curves is by *comparing directly the exposure curves*. After all, the exposure curves (and not the cumulative distribution functions (CDFs)) are the ones that are used to price layers of (re)insurance and assess the impact of different levels of retention. See Appendix B for a more in-depth discussion of why the KS distance between the CDFs is too restrictive a way of comparing distributions for use in exposure rating. The most obvious way of doing this comparison is perhaps by using the root-mean-square (RMS) distance between the exposure curves that we have already used as one of the ways (least-squares regression) in which we calculate the parameters of the MBBEFD exposure curve[4]:

$$d_{\mathrm{RMS}}\left(G_{\mathrm{emp}}, G_{\mathrm{MBBEFD}}\right)=\sqrt{\int_0^1\left(G_{\mathrm{emp}}(u)-G_{\mathrm{MBBEFD}}(u)\right)^2 du}. \tag{2.3}$$

Alternatively, and even more simply, we could calculate the KS distance between the empirically derived exposure curve and the fitted MBBEFD. This is consistent with the approach taken in Riegel (2010, Appendix F). Notice that here the KS distance does not have the statistical interpretation from its standard use, but simply measures whether two exposure curves produce similar results throughout the interval $u \in [0, 1]$:

$$d_{\mathrm{KS}}\left(G_{\mathrm{emp}}, G_{\mathrm{MBBEFD}}\right)=\sup_{u\in[0,1]}\left|G_{\mathrm{emp}}(u)-G_{\mathrm{MBBEFD}}(u)\right|. \tag{2.4}$$

Equations (2.3) and (2.4) relate immediately to the error that one makes in assessing the expected ceded losses to an insurance programme: the expected ceded losses $\mathbb{E}(Z_D)$ above a given retention level D for a property with MPL M can be written as $\mathbb{E}(Z_D)=(1-G(D/M))\times\mathbb{E}(Z)$, where $\mathbb{E}(Z)$ are the expected ground-up losses for the whole property. The error is then[5] $\delta\mathbb{E}(Z_D)=\delta G(D/M)\times\mathbb{E}(Z)$, and the maximum error (in absolute monetary amount) is

$$\sup\left(\delta\mathbb{E}(Z_D)\right)=d_{\mathrm{KS}}\left(G_{\mathrm{emp}}, G_{\mathrm{MBBEFD}}\right)\times\mathbb{E}(Z), \tag{2.5}$$

while the average squared error is

$$\begin{aligned}\mathrm{RMS}\left(\delta\mathbb{E}(Z_D)\right) &= \sqrt{\int_0^1\left(G_{\mathrm{emp}}(u)-G_{\mathrm{MBBEFD}}(u)\right)^2 du}\times\mathbb{E}(Z)= \\ &= d_{\mathrm{RMS}}\left(G_{\mathrm{emp}}, G_{\mathrm{MBBEFD}}\right)\times\mathbb{E}(Z).\end{aligned} \tag{2.6}$$

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

Connection probability	d_{RMS}	d_{KS}	Comments
$p = 0.025$	1.3%	7.9%	d_{KS} is quite high because G is quite steep for low values of u.
$p = 0.05$	1.6%	2.8%	
$p = 0.1$	1.0%	1.8%	This is in practice the most relevant case because losses are of the same order of magnitude as the whole graph – the case that exposure curves are designed to capture.

FIGURE 7: Goodness of fit of the MBBEFD model for different values of the connection probability in the case of complete graphs.

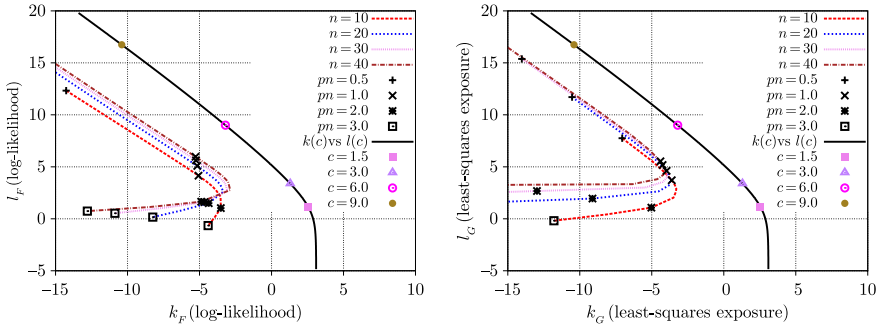


FIGURE 8: Values of k and l corresponding to different values of p and n. All the values of p for a given n lie on one of the dashed lines. The black continuous line shows the values of k and l that correspond to Swiss Re curves with different values of c. The value c= 1.5 (the standard choice for residential houses) has been highlighted. The left and right charts represent two different ways of calculating the parameters, respectively, through MLE and through least-squares fitting of the exposure curve.

Figure 7 shows the RMS distance and the KS distance between the empirical exposure curve and the MBBEFD model for the cases shown in Figure 6: below, roughly at or above the critical threshold $p_c = 1/19$.

Overall, the MBBEFD exposure curves appear to be a good model for the graph-derived exposure curves.

Notice that the curves obtained in this way are not close to Swiss Re's "c curves", a subset of the MBBEFD curves whose use is quite widespread in property insurance and is investigated in Guggisberg (2004). This is made evident in Figure 8, which shows the values of k and l for different combinations of p (propagation probability) and n (number for nodes), and investigated further in Appendix E.

Note that as the value of $p \times n$ evolves from $p \times n \ll 1$ to $p \times n \gg 1$, there is a change of behaviour (a turning point) in all dashed lines, which appears related to the asymptotic behaviour around the critical threshold, as described by Equation (2.1). However, this relationship has not been investigated further (see Section 5.2.6 for further commentary on this).

The practical interest of these initial results, however, is limited as complete graphs do not represent realistic models for properties. We therefore move on to consider a portfolio of actual properties and their graph representation.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

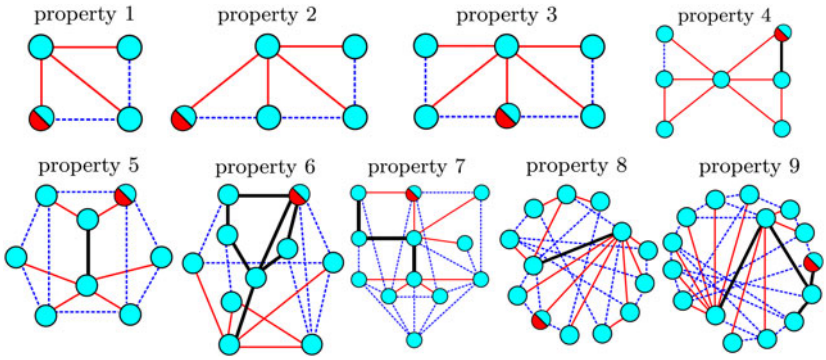


FIGURE 9: The property graphs for a portfolio of different property structures, which reflect typical houses/flats in the UK with varying numbers of bedrooms. The arcs are colour-coded (and also format-coded) depending on the ease of propagation: a red (continuous and thin) arc stands for a door; a blue (dashed) arc stands for a wall or floor/ceiling; a black (continuous and thick) arc stands for an open space. The bi-coloured nodes represent the location of the kitchen; this is for use in Appendix C.

2.4.3. *Random graphs based on a portfolio of actual residential properties.* More relevant results can be obtained if the property graphs reflect the structure of actual properties, being far from complete graphs as the number of nodes increases. Figure 9 shows a collection of possible realistic structures. Property 7 is the example presented at the beginning of Section 2.

Using our standard simulation process we can associate a severity and exposure curve for each of the property structures in Figure 9. These are shown in Figure 10. Note that when doing this we are assuming – as in the case of complete graphs – that each node has a unit value and that the fire can start at any of the nodes with equal probability. These assumptions can be relaxed (see Appendix C) without a substantial change in the results.

Interestingly, we see that the values of pairs of parameters (k, l) evolve (going from simpler to more complex structures) in very much the same way as the pairs corresponding to different values of c. However, the location of these pairs in the parameter space is quite separate from the curve $k = k(c), l = l(c)$, meaning that the empirical severity/exposure curves are not similar to the Swiss Re c curves.

We have so far considered curves for specific properties. However, and especially for the case of residential properties where individual underwriting is impractical, it is also (and perhaps more) important to identify average curves for the whole portfolio. This is also shown in Figure 10.

In order to derive the portfolio curves we have simulated losses by first selecting at random one of the properties in Figure 9 and then removing arcs at random for that property reflecting the propagation probability for rooms separated by a door (P_D), by a wall (P_W) or not separated (P_O). The results in Figure 10 have been obtained by using $P_D = 0.5, P_W = 0.1, P_O = 1$. The probability of selecting a given type of property reflects the percentage of properties in England and Wales with a given number of bedrooms. These are shown in Figure 11.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

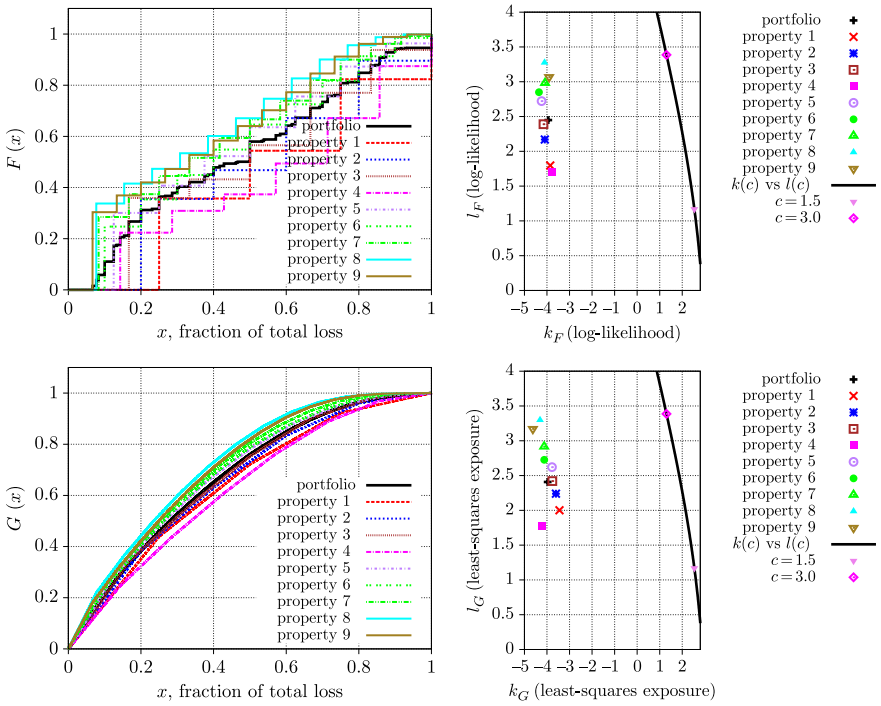


FIGURE 10: (Left) The severity and exposure curves for the property graphs shown in Figure 9, and for the whole portfolio, obtained simulating 100, 000 different scenarios. (Right) The parameters k, l for the various property structures and the portfolio (calculated by maximising log-likelihood, top, and by least squares regression on the exposure curves, bottom), compared with the values of k and l corresponding to different values of c for the Swiss Re c curves.

Figure 12 shows again the empirically derived severity and exposure curve for the whole portfolio, against the MBBEFD curve that fits this curve best. As in the case of the complete graphs, the MBBEF curve provides a good fit to the empirical exposure curves: the RMS and KS distances are 1.0% and 1.5%, respectively.

3. ALTERNATIVE APPROACH: THE RANDOM TIME APPROACH

The random graph approach outlined in Section 2 relies on a "static" view of fire propagation: the fire pathways are defined at random in advance, and the loss is just the sum of the values of the nodes connected to the unit where the fire originates. There is no attempt to capture realistically how the fire actually spreads.

In this section, we introduce a model that still relies on graph-theoretic concepts but views fire propagation as a dynamic process. In this dynamic model, fire propagates through a building deterministically, but for a random time.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

No of bedrooms	% in the portfolio
1	11.9%
2	27.7%
3	41.5%
4	14.0%
≥5	4.9%

Property	No. of bedrooms	No. of rooms (order)	Graph size	Graph density	Probability
1	1	4	5	0.833	0.0596
2	1	5	7	0.700	0.0596
3	2	6	9	0.600	0.1384
4	2	7	10	0.476	0.1384
5	3	8	15	0.536	0.2076
6	3	10	19	0.422	0.2076
7	4	12	27	0.409	0.1888
8	4	13	28	0.359	0.1888
9	5	15	34	0.324	0.0492

FIGURE 11: The portfolio composition, based on (National Archives, 2011), on the percentage of households (both houses and flats) with a given number of bedrooms. So, for example, 11.92% of properties have one bedroom (sum of the first two rows). Note that 4.92% is actually the percentage of households with five bedrooms or more, but, not having graphs for larger buildings available, we used the five-bedroom graph for this building class.

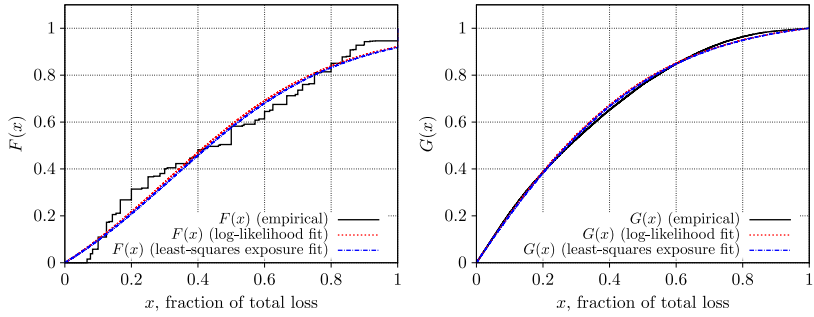


FIGURE 12: The severity and exposure curves for the portfolio (Figure 9), compared with the best-fit MBBEFD curve. The best-fit curve is again derived in two different ways: by an MLE fit of the severity distribution and by a least-squares fit of the exposure curve.

As in the model considered in Section 2, nodes correspond to rooms (or property units) with a fixed monetary value, and arcs connect rooms between which fire can propagate. Weights distinguish physical situations (open door, closed door, wall) and are related to the time it takes for fire to spread.

The fire will spread for a time T (which we will refer to as "delay"), which depends on how long it takes

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

(a) for sprinklers/extinguishers to be activated;
(b) for the fire to be reported;
(c) for the emergency services to arrive on site[6]; and
(d) for them to put the fire out.

The shape of the (overall) delay distribution should ideally be based on available statistics, but it is unrealistic to expect that this level of information is available except in specific, well-studied cases. An ansatz based on a simple exponential distribution (as is quite common for waiting times) for the delay time is probably all we can aim for in most situations. For the experiments in this paper we have considered a simple exponential distribution for the time it takes before the fire is extinguished:

$$F_T(t) = 1 - \exp(-\lambda t), \tag{3.1}$$

where the parameter λ is inversely related to the expected time:

$$\mathbb{E}(t) = \frac{1}{\lambda} = 2 \text{ time steps} \sim 6 \text{ min}. \tag{3.2}$$

Also note that in this model, unlike the model described in Section 2 and similarly to one of the model refinements addressed in Appendix C, we allow for different values of the rooms.

The simulation of fire propagation proceeds as follows:

1. Generate a time over which the fire will propagate (includes fractions of steps).
2. Select a property from the portfolio, according to the available statistics on the number of bedrooms (National Archives, 2011).
3. Choose a monetary value for each room (a random number uniformly distributed between 0 and 1)[7].
4. Set the doors open/closed with probability $P(\text{open})$.
5. Choose a room at random where the fire starts.
6. Propagate – fire traverses each edge from a burnt room in:
   a. Open space/open door $\rightarrow n_{\text{open}}$ time steps.
   b. Closed door $\rightarrow n_{\text{closed}}$ time steps.
   c. Wall/ceiling/floor$\rightarrow n_{\text{wall}}$ time steps.
7. Once done, report damage ratio.
8. Repeat 100,000 times (Monte Carlo) to get empirical CDF.

The results of the simulation with a given set of parameters are shown in Figure 13.

As in the case of the random graph approach, the MBBEFD model gives a good approximation of the various curves coming out of the simulation (the RMS distance and KS distance between the exposure curves are 0.8% and 1.4%, respectively), for both the individual properties and the portfolio, but the values of the parameters are far from those of the Swiss Re c curves (see also

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

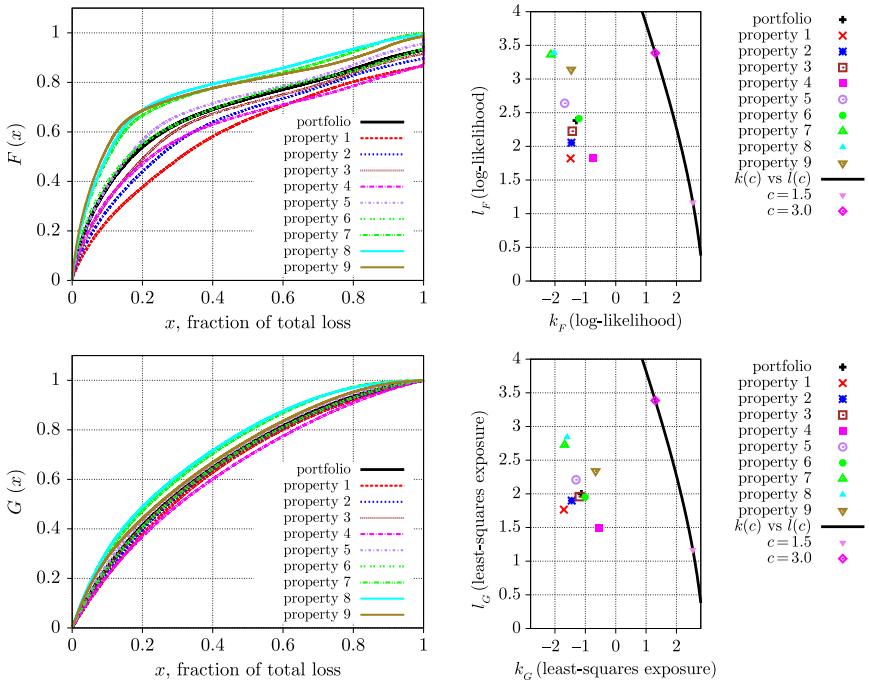


FIGURE 13: The result of the simulation with the random time approach with the following parameters: $\lambda = \frac{1}{2}$, P (open) $= 0.8$, $n_{\text{open}} = 1$, $n_{\text{closed}} = 10$, $n_{\text{wall}} = 30$. (Left) The severity (top) and exposure (bottom) curves for the different property structures of Figure 8, and for the whole portfolio, obtained simulating 100,000 different scenarios. (Right) The parameters k, l of the MBBEFD curves for the various property structures and the portfolio, calculated again according to two different methods (maximum likelihood and least-square regression) compared with the values of k and l corresponding to different values of c for the Swiss Re c curves (the black curve).

Appendix E). This seems to be true also for a wide range of different choices of the parameters.

4. MODELLING EXPLOSION RISK

The random graph approach can be extended to other perils beyond fire, such as explosion. Explosion risks are typically assessed with proprietary tools such as Swiss Re's ExTool[8]. These work by assessing the damage of an explosion for all units that may be the site of the explosion, and estimate the MPL as the maximum of the loss over all units.

An explosion may be due, for example, to the formation of a vapour cloud following a leak of flammable liquid or to the failure of pressure vessels. Once the explosion occurs, the damage extends to all property units within a certain range, and the degree of damage decreases with the distance from the explosion. ExTool has three levels of damage: 80%, 40% and 5%. The picture may be complicated by secondary fires, considerations around debris removal and firefighting expenses.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

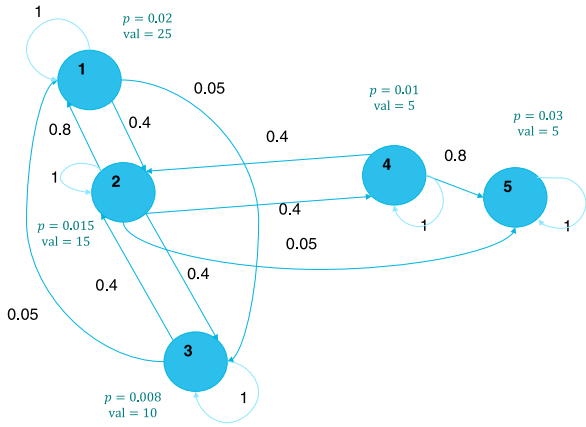


FIGURE 14: Explosion graph for a simple plant.

In some respects, modelling explosions with graphs is easier than for fire, as, unlike wide-spreading fire, the danger of a (large) secondary explosion is negligible[9]. The number of nodes affected by the explosion in a node is therefore simply determined by the physical/chemical features of each unit and the distance to the other units. In other respects, explosions are more complex as it requires to take directionality into account: the effect on B of an explosion in A is not necessarily the same as the effect on A of an explosion in B (e.g., A might cause a bigger explosion implying total destruction of B but not vice versa).

We therefore need to use a directed graph, or digraph, to model this situation. An example of how an explosion graph looks like is given in Figure 14, which represents a petrochemical plant with 5 units, each of which can be the site of an explosion.

Each of the nodes in the graph in Figure 14 represents a unit. In this particular example, each site can be the site of an explosion (that is not always the case). There is an arc from node A to node B if an explosion in site A damages site B. For this reason, there is also an arc between A and A itself with weight 1 (total loss) – these self-node arcs are drawn in a lighter colour to avoid clutter but they have exactly the same status as the other. The weight of the arc from A to B is the percentage of damage to B. For example, node 2 affects all other nodes but causing different damage ratios. The weight from B to A does not need to be equal to the weight from A to B: for example, the arc from 1 to 2 is only 0.4 (while from 2 to 1 is 0.8) – this is possible because of the different processes taking place in unit 1.

In more general terms, the di-graph associated with a property can be constructed as follows:

(a) A value is associated to every unit v, val(v). This is the maximum loss should unit v explode.
(b) Each unit v has an associated probability $p(v)$ of originating an explosion. $p(v)$ may be 0 for some or even most of the units.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

(c) From each of these units, *directed arcs* (i.e., arcs with a direction) can be drawn to all the units that are within reach of it should an explosion occur. For example, unit v is connected with units $u_1, u_2 \ldots u_N$ through arcs $(v, u_1), (v, u_2) \ldots (v, u_N)$. Note that for a directed arc the direction matters: (v, u_1) and (u_1, v) are two different arcs.

(d) A weight $w = w(v, u)$ can be assigned to each arc (v, u), giving the proportion of loss that we expect to incur to unit u as a result of an explosion in unit v. The weight will be a function of the distance, as modelled by explosion tools. Note that in general $w(v, u) \neq w(u, v)$.

(e) The loss consequential to an explosion in unit v (which is assumed to be fully destroyed as a consequence of it) can be calculated as

$$\text{Loss}(v) = \text{val}(v) + \sum_{j=1}^{N} w_j \times \text{val}\left(u_j\right), \tag{4.1}$$

where w_j is shorthand for $w\left(v, u_j\right)$.

(f) The MPL can then be defined as the maximum of Loss (v) as calculated over all possible originating sites v (i.e., all sites for which $p_v > 0$):

$$\text{MPL} = \max\left\{\text{Loss}(v) \,|p_v > 0\right\}. \tag{4.2}$$

Based on this representation, the calculation of a severity curve is straightforward and does not need Monte Carlo simulation (although it can also be approximated in that way). The process is explained in the algorithm below and illustrated in the example in Section 4.1.

Severity/exposure curve (explosion) – algorithm

Input: a graph $G = (V, M, P, A, W)$ where $V = (v_1, v_2, \ldots v_N)$ is the set of nodes, $M = (\text{val}(v_1), \text{val}(v_2) \ldots \text{val}(v_N))$ is the set of values associated with each node, $P = (p(v_1), p(v_2), \ldots p(v_N))$ is the set of probabilities associated with each node, A is the set of (directed) arcs and W is the set of the weights associated with each arc:

1. For each unit v calculate the Loss(v) should an explosion start at that site, as per Equation (4.1) (the units for which $p(v) = 0$ can be omitted).
2. Calculate MPL as the maximum of Loss(v) across all v (Equation (4.2)).
3. For each unit v calculate the damage ratio, Frac_Loss(v) as $\text{Frac_Loss}(v) = \text{Loss}(v)/\text{MPL}$.
4. Sort the damage ratios in ascending order: $x_1 < x_2 < \cdots < x_N = 1$ (without loss of generality, we can consider

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

them all distinct - see next step for the treatment of identical damage ratios).

5. Associate to $x_1 < x_2 < \cdots < x_N$ the probability of explosion of the corresponding units. If two or more units $v, v', \ldots$ have the same damage ratio x_j, the sum of the relevant probabilities is used and k is decreased accordingly. The output of this step is N probabilities $p_1, \ldots p_N$.
6. Calculate the proportion $\pi_1, \ldots \pi_N$ of the damage ratios $x_1, \ldots x_N$ by normalising the probabilities in Step 5 to unity: $\pi_j = p_j / \sum_i p_i$.
7. The CDF $F(x)$ can then be calculated as follows:

$$F(x) = \begin{cases} 0 & 0 \le x < x_1 \\ \sum_{i=1}^{j} \pi_i & x_j \le x < x_{j+1} \\ 1 & x = 1 \end{cases}$$

for $j = 1, \ldots N-1$.

8. The exposure curve $G(d)$ can be calculated as usual as the normalised integral of the survival function $S(x) = 1 - F(x)$:

$$G(d) = \frac{\int_0^d S(x)\, dx}{\int_0^1 S(x)\, dx.}$$

(This integral is trivial to calculate for a piecewise-constant function as in our case.)

A similar result can be obtained by Monte Carlo simulation, which is very straightforward since the only stochastic element (in the basic version of the algorithm) is the selection of a particular site for the explosion. The process can be carried out by replacing Steps 4–8 in the algorithm above with the following pseudocode:

Simulation process – algorithm

Input and Steps 1-4 as above

For all scenarios $j = 1, \ldots N$.

Pick a node $v \in V$ (explosion site) at random with probability $\frac{p(v)}{\sum_{v'} p(v')}$.

Loss for scenario j: $\text{Loss}(v) = \text{val}(v) + \sum_{j=1}^{k} w_j \times \text{val}(u_j)$, where u_j $(j = 1, \ldots k)$ are the nodes connected to v via a directed arc starting in v.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

Value		25	15	10	5	5		
Prob		**1**	**2**	**3**	**4**	**5**	Loss	Damage ratio
0.020	**1**	1.00	0.40	0.05	0.00	0.00	31.5	76.4%
0.015	**2**	0.80	1.00	0.40	0.40	0.05	41.3	100.0%
0.008	**3**	0.05	0.40	1.00	0.00	0.00	17.3	41.8%
0.010	**4**	0.00	0.05	0.00	1.00	0.80	9.8	23.6%
0.030	**5**	0.00	0.05	0.00	0.00	1.00	5.8	13.9%

FIGURE 15: Matrix representation of an explosion di-graph.

Damage ratio for scenario j: $\text{Damage_Ratio}(v) = \text{Loss}(v) / \text{MPL}$.

Sort damage ratios in ascending order to produce a normalised severity curve and calculate the corresponding exposure curve.

Obviously, the severity and exposure curves calculated in this way are just an approximation of the ones for the actual physical explosion process. The damage percentage based on the distance from the explosion does not, in reality, vary discretely, and the range of the explosion will not have an entirely predictable shape and size. What we have proposed here is a method to build approximate severity/exposure curves that are just as sophisticated as a typical simple output from an explosion modelling tool.

As for the fire propagation graphs described in Section 2, matrices provide an alternative useful representation for the explosion di-graphs. For example, the di-graph in Figure 14 can be represented by the matrix in Figure 15. Note that this matrix is in general non-symmetrical.

The way to read the matrix in Figure 15 is as follows: the numbers in bold are the unit IDs. The rows represent the starting point of the arc, the columns the endpoint, and the figure at the intersection is the weight. For example, there is an arc between unit 2 and 1 whose weight is 0.8, while the arc between 1 and 2 has weight 0.4. The loss associated to a particular unit can then be calculated as the scalar product of the value vector (25, 15, 10, 5, 5) and the damage percentage vector (e.g., (1, 0.4, 0.05, 0, 0) for unit 1). By dividing the loss of a unit by the MPL (the maximum of losses for units 1–5) one obtains the damage ratio, which is used to build the normalised severity curve. The first column gives the probabilities that an explosion starts at a particular site.

This example shows that based on the adjacency matrix, it is straightforward to calculate the loss associated with each unit, and from that the losses as a percentage of MPL. This allows us to build the normalised severity as a piecewise constant function where the steps occur at the different damage ratios for each unit and the height of each step is proportional to the probability of explosion for that unit (see Figure 16).

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

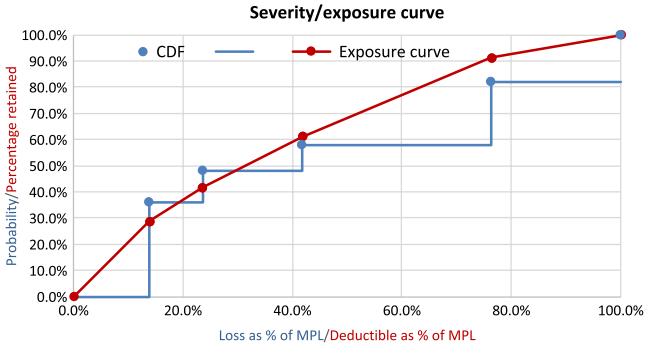


FIGURE 16: Severity and exposure curves associated with an explosion graph. The severity curve $F(x)$ for the explosion graph in Figure 14 is a stepwise function, where the steps occur at the damage ratios shown in Figure 15. The corresponding exposure curve $G(d)$ is a piecewise linear function.

5. WRAPPING UP

5.1. Conclusions

We have produced severity and exposure curves for fire from first principles, based on a graph representation of properties, using two paradigms – the random graph approach and the random time approach.

For both paradigms, the MBBEFD framework turns out to be rich enough to provide a good approximation for the exposure curves arising from these graphs. However, the commonly used Swiss Re c curves do not tie up with the behaviour we see here. This means that either the graph-theoretic view is not (yet) realistic enough or that the c curves are too restrictive.

The random graph approach can be extended to model explosion risk. For this case we need to use a more refined graph model (directed graphs), but in its basic incarnation the mathematical treatment is much simpler than for fire.

5.2. Limitations and future research

The idea of using graph-theoretic techniques to bridge the gap between the physical description of a property and the statistical distribution of losses is rich of potential. In this initial study we have only scratched the surface of what could be done with it. There are many directions in which this study could be developed.

5.2.1. *A more realistic graph representation.*
In the random graph model for fire, each point represents a total loss of a unit, and the value of each unit is the same. This appears to be unduly restrictive: (a) different nodes could have different values, (b) different nodes could have different propensities to start a fire, and (c) the loss per unit could be stochastic (i.e., a number between 0 and the value of that unit sampled from a statistical distribution). Appendix C discusses the effect of extending the model

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

so as to incorporate (a) and (b). As for (c), initial calculations show that making the loss to a node stochastic has the unintended consequence of making the probability of a total loss negligible (which is unrealistic), unless the loss distribution associated with the node is complex enough to accommodate a substantial probability of a total loss for that node. Such additional complexity would arguably obfuscate rather than elucidate the relationship between graph constructs and exposure curves. A more promising avenue for taking into account the possibility of a partial loss for a node is perhaps to split each room into a number of smaller units, e.g. a cooking appliance might be a unit instead of the kitchen. The connectivity within a single room would of course be stronger than between units in different rooms.

More in general, the results of professional fire propagation tools and specialist studies in fire propagation could be used to calibrate property graphs. This is true for both the random graph and the random time approach. It is an open question whether the added complexity of using these models for the purpose of building exposure curves for portfolios of similar properties (rather than for the purpose for which they were originally built) will lead to improved severity curves or simply spurious accuracy.

In the graph model for explosions, the models can be made more realistic by dispensing with the discretisation in the damage percentage of the affected units and introducing stochastic elements in both the percentage damage and the direction of propagation.

Although all these extensions are possible, it is dubious whether they are all warranted: a simple model is normally preferable to a model which looks more "realistic" but is often simply spuriously accurate and over-parameterised, unless the added complexity is built on the back of significant data.

5.2.2. *Systematic study of models and parameterisations.*
A systematic study of the relationship between the model parameters (e.g., the weight of the arcs between nodes) and the shape and parameters of the exposure curves has not been attempted, although some trends have been identified (e.g., see discussion in Section 2.4.3 on the evolution of the parameters k, l as graphs become more complex). Such study would enable us to respond to questions such as: Is it possible to approximate efficiently the MBBEFD curves for the random graphs model (Section 2.4) and the random time model (Section 3) with a single-parameter model similar to Swiss Re's c curves? The way in which best-fit values of the parameters k and l arrange themselves as a curve on the (k, l)-plane (see Figures 10 and 13) appears to suggest that there may be an underlying parameter γ in terms of which k and l can be conveniently expressed: $k = k(\gamma)$, $l = l(\gamma)$. Whether that is indeed the case, and whether different curves would be needed for different values of the arc weights, requires further research.

The extent to which the MBBEFD model itself can be improved upon as an approximate model for graph-based exposure curves has also not been explored and more research is needed. Examples of alternative models are

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

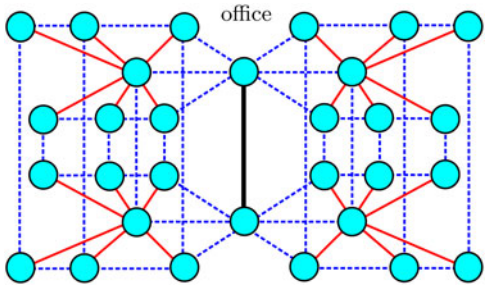


FIGURE 17: The graph representation of a commercial office.

the Beta distribution with the addition of a point mass of given probability at $x = 1$ equal to the probability of a total loss, or models that may arise as natural approximations in the context of critical phenomena theory (see also Section 5.2.6. for further discussion).

5.2.3. *Extension to other types of property.*
The focus of this study has mainly been residential properties, because they were homogeneous enough as a class and because a fair amount of public data on different property structures was available. The methodology can obviously be extended to commercial properties. For some of them – such as offices – the translation of the methodology is straightforward, and the graphs are very much the same as for residential buildings, but more complex because of the larger number of rooms (see Figure 17).

For other, more complex types of commercial properties such as manufacturing plants, refineries, and so on, the approach is still valid but the concept of "property unit" may need to be adapted and does not necessarily correspond to that of "room". For example, for the case of a tank farm, the natural units will be the individual tanks.

5.2.4. *Extension to other perils.*
The methodology could be generalised from fire and explosion to all F/L/Ex/A (fire, lightning, explosion/aircraft collision) losses. We expect that lightning and aircraft collision losses will be similar to explosion losses, in the sense that they have a natural centre (e.g., the point where an aircraft crashes) and there is going to be an affected area around it. However, it is not obvious that we need di-graphs to represent these risks – weighted undirected graphs may be sufficient.

5.2.5. *Extension to other lines of business.*
This investigation has been mainly about property damage risk, which is about physical structures that can normally be segmented into units in a natural away. That makes it amenable to graph representation and modelling.

However, the use of graph-theoretic concepts could be explored in other lines of business where there are interconnected elements, for example, in credit

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

risk or even in liability. One complication for liability is of course that there is no such thing as a "total loss" and the graphs themselves may not be limited.

5.2.6. *Relationship with the theory of critical phenomena.*
The theory of critical phenomena is the natural theoretical framework for studying the size of the connected component and the probability of a total loss. Some useful results have already been mentioned (see Section 2.3). Undoubtedly the literature on this subject is quite rich and there may be more results out there that we could use off-the-shelf. Specifically, we are not aware of results on the distribution of the size of the connected component of a random graph – these would be directly relevant to our investigation.

The main problem is of course that many of these results are of an asymptotic nature, whereas we are normally interested in graphs of very limited cardinality (number of nodes $\sim$ 5–100). Another problem is that these results may need considerable adaptation to the specifics of our situation which is quite complex: nodes may have different values, and weights may be different between arcs.

Acknowledgements

We would like to thank Eric Lenoir and Jean-Christophe Candelon for very helpful discussion on the technical aspects of explosion risk and fire risk. We are also deeply grateful to one of the anonymous referees for their helpful suggestions on the question of the correct metrics for judging the adequacy of the MBBEFD model: the content of Appendix C was largely inspired by such suggestions.

Notes

1. https://www.nist.gov/publications.
2. http://www.swissre.com/clients/client_tools/about_extool.html.
3. In this paper we often speak of "empirical" curves not only to denote curves based on actual data but also to denote curves that have been derived as the output of stochastic simulation rather than analytically.
4. It might appear statistically unsound to use the RMS distance as a measure of fit when we have used it to derive the parameters for the exposure curve. However, we are not trying to assess whether MBBEFD is the correct distribution to describe the empirically derived distribution (we know that it is not the case) nor we are attempting to derive a predictive MBBEFD model. All we need to ascertain is if there are parameters for which the MBBEFD model provides a decent interpolation of the empirical curve.
5. The "δ" symbol represents here an absolute difference in amounts, for example, $\delta G(d) = \left|G_{\text{emp}}(d) - G_{\text{MBBEFD}}(d)\right|$, as sometimes used in error theory.
6. This is on average just below 9 min in England.
7. If the graph corresponding to the property is connected, the MPL can be defined as the sum of all values of the nodes.
8. http://www.swissre.com/clients/client_tools/about_extool.html.
9. In the case of vapour cloud explosions, the primary explosion will cause fires to spread to the affected areas. As a consequence, any vapour that is spilled as a consequence of the first explosion is typically set on fire before it accumulates to the point of explosion. Thus, secondary explosions are unlikely. Small explosions are possible but they do not have a significant impact

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

on the overall damage (personal communication by E. Lenoir, risk control practice leader in the Energy sector).

References

Bernegger, S. (1997) The Swiss Re exposure curves and the MBBEFD distribution class. *ASTIN Bulletin*, **27**(1), 99–111.

Bollobas, B. (2011) *Random Graphs*. Cambridge: Cambridge University Press.

Brehm, J.J. and Mullin, W.J. (1989) *Introduction to the Structure of Matter*. New York: Wiley.

Center for Chemical Process Safety. (2010) *Layer of Protection Analysis: Simplified Process Risk Assessment*. New York: Wiley.

Chhabra, A. and Parodi, P. (2010) Dealing with sparse data. *Proceedings of GIRO 2010*, Newport.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.

Freund, J. E. (1999) *Mathematical Statistics*. 6th Edition. New Jersey: Prentice Hall International, Inc.

Frieze, A. (2015) *Introduction to Random Graphs*. Cambridge: Cambridge University Press.

Guggisberg, D. (2004) Exposure Rating. Zurich: Swiss Re. http://www.swissre.com.

Hagberg, A., Schult, D., Swart, P. (2018) NetworkX reference, release 2.1, https://networkx.github.io/documentation/stable/_downloads/networkx_reference.pdf.

Kraft, D. (1988) A software package for sequential quadratic programming. Tech. Rep. DFVLR-FB 88-28, Köln, Germany: DLR German Aerospace Center – Institute for Flight Mechanics.

Mandelbrot, B. B. (1982) *The Fractal Geometry of Nature*. San Francisco: W.H. Freeman and Co.

National Archives. (2011) Office For National Statistics. Part of 2011 Census. Detailed Characteristics on Housing for Local Authorities in England and Wales Release. Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.3.0. Crown Copyright 2015. http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census/detailed-characteristics-on-housing-for-local-authorities-in-england-and-wales/sty-households-in-england-and-wales.html.

Pareto, V. (1964) *Cours d'Économie Politique: Nouvelle édition par G.-H. Bousquet et G. Busino,* pp. 299–345. Geneva: Librairie Droz.

Parodi, P. (2014) *Pricing in General Insurance*. New York: CRC Press.

Rath, B. and Toth, B. (2009) Erdos–Renyi random graphs + forest fires = self-organized criticality. *Electronic Journal of Probability* , **14**, Paper no. 45, 1290–1327. doi:10.1214/EJP.v14-653. https://projecteuclid.org/euclid.ejp/1464819506.

Reason, J. (1990) The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **327**(1241), 475–484.

Riegel, U. (2010) On fire exposure rating and the impact of the risk profile type. *ASTIN Bulletin*, **40**(2), 727–777.

Slade, G. (2008) Probabilistic models of critical phenomena. In *The Princeton Companion to Mathematics* (ed. Gowers et al.), pp. 657–670. Princeton: Princeton University Press.

Trudeau, R.J. (2003) *Introduction to Graph Theory*. New York: Dover Publications Inc.

UK Government. (2010) The building regulations 2010. Fire safety. *Approved Document B*. Vol. 1.—Dwelling Houses: Crown Copyright (2011). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/485420/BR_PDF_AD_B1_2013.pdf.

Willey, R.J. (2014) Layer of protection analysis. *Procedia Engineering*, **84**, 12–22. https://www.sciencedirect.com

Pietro Parodi (Corresponding author)
SCOR Global P&C
London, UK
E-Mail: pparodi@scor.com

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

PETER WATSON
SCOR Global P&C
London, UK
E-Mail: pwatson@scor.com

APPENDICES

APPENDIX A. EXPOSURE CURVES

Property insurance is typically priced (fully or in part) via exposure curves. These can be seen as re-engineered versions of severity curves that allow to assess the effect on the expected losses of changing the retention level, or allocate the expected losses to layers of (re)insurance. The main definitions are below. Refer to Parodi (2014), Chapter 21, for a more detailed treatment.

Assume that X represents a loss to a particular property, and that the maximum possible loss (MPL) for that property is M. Since the loss cannot exceed M, it is quite natural to describe any loss occurring to that property as a percentage of its maximum value M, $x = X/M$ (damage ratio, also referred to in the literature as normalised loss, loss degree, degree of loss, fractional loss.). Also, we can describe the deductible D as a percentage of M, $d = D/M$.

The exposure curve is defined by the following function $G(d)$, which gives the percentage of the expected losses retained by the (re)insured if a deductible of value $d \times M$ is imposed.

$$G(d) = \frac{\int_0^d (1 - F(x))\, dx}{\int_0^1 (1 - F(x))\, dx} = \frac{\int_0^d S(x) dx}{\int_0^1 S(x) dx} \tag{A.1}$$

In Equation (A.1), $F(x)$ is the normalised severity curve (or normalised CDF), i.e. the probability that the damage ratio is less or equal to x, and $S(x) = 1 - F(x)$ is the survival function.

As mentioned in Section 1, the exposure curve is in 1:1 correspondence with $F(x)$. Equation (A.1) shows how to obtain G from F. Equation (A.2) shows how to obtain F from G:

$$F(x) = \begin{cases} 1 - \frac{G'(x)}{G'(0)} & 0 \le x < 1 \\ 1 & x = 1 \end{cases} \tag{A.2}$$

The MBBEFD class of distributions is defined as

$$G(d) = \frac{\ln\left(\frac{(g-1)b + (1-bg)b^d}{1-b}\right)}{\ln(bg)} \tag{A.3}$$

$$F(x) = \begin{cases} \frac{b(g-1)(1-b^x)}{b(g-1) + (1-bg)b^x} & \text{if } x < 1 \\ 1 & \text{if } x = 1 \end{cases} \tag{A.4}$$

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

The Swiss Re "c" curves are a special case of the MBBEFD class, with b and g calculated as

$$b = e^{\alpha+\beta c(1+c)}, g = e^{(\gamma+\delta c)c} \tag{A.5}$$

with $\alpha = 3.1$, $\beta = -0.15$, $\gamma = 0.78$, $\delta = 0.12$

Clearly, Equation (A.5) defines a parametric curve in the (b, g) plane.

In this paper, we have used a slightly different parameterisation than the usual (b, g) one. We have used parameters (k, l) that are related to (b, g) by these equations:

$$b = e^k,\ g = 1 + e^l \tag{A.6}$$

The rationale behind this parameterisation is explained in Section 2.4.1.

APPENDIX B. ON ASSESSING THE ADEQUACY OF EXPOSURE CURVE MODELS

To assess the adequacy of MBBEFD curves for describing the exposure curves arising from our graph-based models we have used the KS distance and the RMS distance applied directly on the exposure curves, rather than, e.g., the KS statistic applied to the underlying severity curves. The rationale behind it is twofold:

(a) The exposure curve is related straightforwardly to the premium charged for a particular layer of (re)insurance, and therefore it is the most relevant object from a business point of view;
(b) Closeness of the severity curves is not a good proxy for the closeness of the relevant exposure curves. This is because if F and F^* are close, the relevant exposure curves G and G^* will also be close, but the converse is not true.

Point (a) requires no further elaboration. As for Point (b), let's first note that if the KS distance between two severity curves F, F^* is less than ε, the KS distance between G and G^* will be bound by this formula (proof below):

$$d_{KS}\left(G, G^*\right) \leq \frac{2\varepsilon}{\int_0^1 S\left(x\right) dx} + O\left(\varepsilon^2\right) \tag{B.1}$$

In other terms, the distance between G and G^* can be made arbitrarily small by making sure that the distance between F and F^* is small enough.

The proof is straightforward and hinges on the fact that one can write the survival distribution $S^*(x)$ as $S^*\left(x\right) = S(x) + \varepsilon D(x)$ where $-1 \leq D\left(x\right) \leq 1$ for all x. Therefore,

$$\left|G^*\left(d\right) - G(d)\right| = \left|\frac{\int_0^d S\left(x\right) dx + \varepsilon \int_0^d D\left(x\right) dx}{\int_0^1 S\left(x\right) dx + \varepsilon \int_0^1 D\left(x\right) dx} - \frac{\int_0^d S\left(x\right) dx}{\int_0^1 S\left(x\right) dx}\right|$$

$$= \left|\frac{\varepsilon\left(\int_0^d D\left(x\right) dx \int_0^1 S\left(x\right) dx - \int_0^d S\left(x\right) dx \int_0^1 D\left(x\right) dx\right)}{\left(\int_0^1 S\left(x\right) dx + \varepsilon \int_0^1 D\left(x\right) dx\right) \int_0^1 S\left(x\right) dx}\right|$$

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

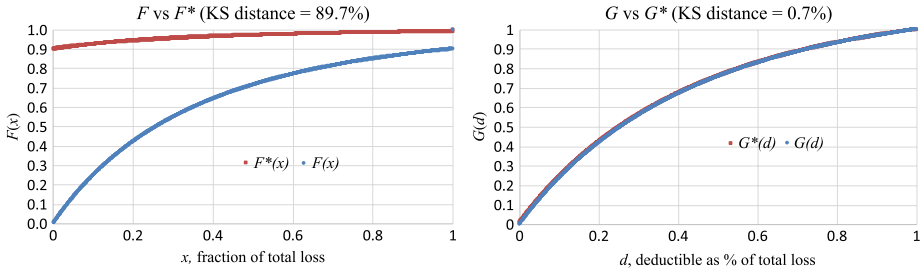


FIGURE B1: The severity curve of $F(x)$ is an MBBEFD curve with $b = 0.15$, $g = 10$. The severity curve of $F^*(x)$ is obtained from $F(x)$ using Equation (B.2) with $k = 0.9$, $\varepsilon = 0.0003$. The KS distance between the severity curves is 89.7%, while the KS distance between the relevant exposure curves is only 0.7% (the difference is barely visible in the graph).

Using the inequalities $\int_0^d S(x)\,dx \le \int_0^1 S(x)\,dx$ and $\int_0^d D(x)\,dx \le 1$, which hold for all values of $d \in [0, 1]$, we can easily derive that $|G^*(d) - G(d)| \le \frac{2\varepsilon}{\int_0^1 S(x)dx} + O(\varepsilon^2)$ for all values of $d \in [0, 1]$ and therefore for the supremum over [0, 1], yielding Equation (B.1).

The converse, however, is not true, and we can show it with a simple counter example. Consider a severity curve $F(x)$, and another severity curve $F^*(x)$ built as follows, for a fixed k, $0 < k < 1$:

$$F^*(x) = \begin{cases} (1-k)\,F(x) & 0 \le x < \varepsilon \\ k + (1-k)\,F(x) & \varepsilon \le x \le 1 \end{cases} \tag{B.2}$$

where $0 < \varepsilon < 1$. A lower bound for the KS distance between F and F^* is then given by:

$$d_{KS}(F, F^*) = \sup_{x\in[0,1]} |F^*(x) - F(x)| \ge |F^*(\varepsilon) - F(\varepsilon)| = k(1 - F(\varepsilon))$$

Assuming that F maps [0,1] into [0,1] and is continuous over [0,1) (as, e.g., the MBBEFD curves are), we can choose ε small enough that $F(\varepsilon)$ is below an arbitrarily small number h, so that $d_{KS}(F, F^*)$ is arbitrarily close to k:

$$d_{KS}(F, F^*) \ge k(1 - h) \tag{B.3}$$

If $G(d)$ and $G^*(d)$ are the exposure curves for $F(x)$ and $F^*(x)$ respectively, the absolute difference between the exposure curves (for $d \ge \varepsilon$) is:

$$\begin{aligned} |G^*(d) - G(d)| &= \left| \frac{(1-k)\int_0^d S(x)\,dx + k\varepsilon}{(1-k)\int_0^1 S(x)\,dx + k\varepsilon} - G(d) \right| = \left| \frac{k\varepsilon}{(1-k)\int_0^1 S(x)\,dx + k\varepsilon}(1 - G(d)) \right| \\ &\le \frac{k\varepsilon}{(1-k)\int_0^1 S(x)\,dx + k\varepsilon} \end{aligned} \tag{B.4}$$

Inequality (B4) also holds for $d < \varepsilon$. And since the right-hand side of the inequality is independent of d, it is also satisfied by the supremum over [0, 1]:

$$d_{KS}(G, G^*) \le \frac{k\varepsilon}{(1-k)\int_0^1 S(x)\,dx + k\varepsilon} \tag{B.5}$$

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

By choosing k sufficiently large and then ε sufficiently small, we can then make $d_{KS}(F, F^*)$ as close to 1 as desired and – once the value of $d_{KS}(F, F^*)$ is chosen – we can make $d_{KS}(G, G^*)$ as close to 0 as desired.

Figure B1 provides an example of two severity curves that are widely separated but however give rise to very similar exposure curves.

The "physical" interpretation of this phenomenon is straightforward. The curve $F^*(x)$ is derived from the curve $F(x)$ by mixing $100k\%$ attritional losses (all of size ε) with $100(1-k)\%$ of losses coming from $F(x)$. As long as the total of these attritional losses is a small percentage of the overall losses, the impact on the exposure curve will be minimal, regardless of how large their frequency is.

The reasoning above and the example in Figure B1 should bring into clear relief what we stated at the beginning: closeness of the exposure curves (and not of the severity curves) is what matters when assessing the adequacy of specific models such as the MBBEFD.

Notice that similar observations would hold for the RMS distance – however, the calculations are more complicated and we have omitted them.

APPENDIX C. MODEL REFINEMENTS

In Sections 2 and 3 we have shown empirical results for the exposure curve related to graphs with a very simple structure – each node having the same unit value (Section 2 only) and the same probability of starting the fire. This is of course unrealistic – different rooms in a property will have in general different values, and it is known that about 50% of the fires in a house start in the kitchen. In this section we show what happens when our initial hypotheses are relaxed to make the model more realistic.

One way of refining the model is to allow the value of each node to be random. Figure C1 (left) shows the case where in a given simulation the value of each room is a random number uniformly distributed between 0 and 1. This value is then reset for every simulation. We can refine this model further to avoid that the ratio of values between two different rooms is limited: Figure C1 (right) shows the case where the value of each room is a random value uniformly distributed between 0.5 and 1. As a consequence of this, the severity curve $F(x)$ becomes continuous for all points $0 \leq x < 1$. The most significant observation, however, is that the parameters of the resulting exposure curve for the portfolio change little from the base case (unit value for all nodes).

Another way of refining the model is recognising that different nodes will have different probabilities of being the origin of the fire. According to UK Home Office statistics (National Archives, 2011), about 50% of fire incidents in houses start because of cooking appliances. We have therefore run a simulation in which the probability of a fire starting in a specific room in a house with n rooms is $50\% + 50\%/n$ for the kitchen and $50\%/n$ for all the other rooms (note that the "$50\%/n$" term in the kitchen is to account for the possibility that a fire starts in the kitchen for reasons unrelated to the cooking appliances). The results for this case are shown in Figure C2. For comparison purposes, the same figure also shows the results for the extreme case where *all* the fires start in the kitchen.

The most striking feature of these results is that it makes very little difference to the exposure curve where the fire starts, as long as the unit where the fire starts has typical connectivity with the rest of the property: and indeed, the kitchen is rather well connected. (The results are quite different if, for example, we choose the loft as the starting point!)

The difference between the different cases can be shown quantitatively: Figure C3 shows the KS distance and the RMS distance between the empirical exposure curves for the

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

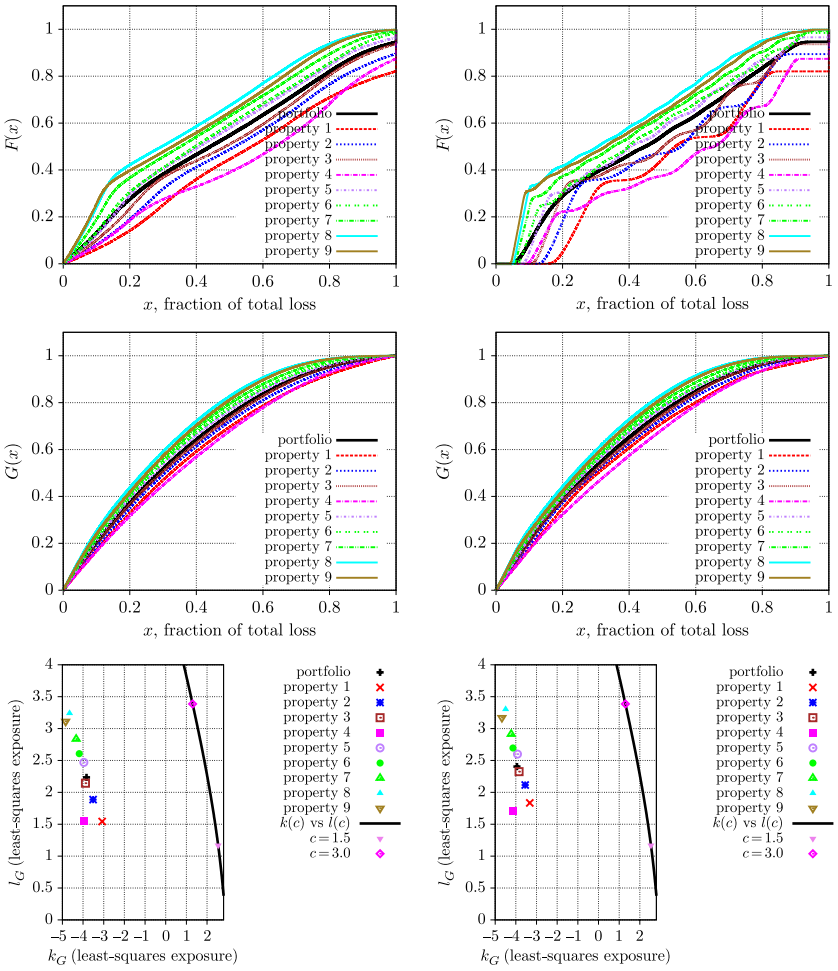


FIGURE C1: (Left) Empirical severity curve (top), empirical exposure curve (middle) and MBBEFD parameter values (bottom) for the case where the value of each node is a random number between 0 and 1. (Right) Same charts for the case where the value of each node is a random number between 0.5 and 1. In both cases, the charts show the curves and values for the individual properties and the whole portfolio as defined in Section 2.4.3.

portfolio in all these different cases. In all cases the difference is smaller or equal than 1.3%, which is within the error due to simulation uncertainty and therefore negligible.

The main lesson to be learnt from the analysis in this section is that sensible refinements to the model lead to results that are very much in line with the simple case where each node has a unit value and each node has the same probability of being the origin of the fire. This suggests that what really matters is the topology of the graph, i.e. the number of nodes and the way they are connected to one another, whilst there is some robustness against changes in the node parameters, i.e. the value associated with each node and the relative probability of starting the fire. This is especially true if we consider not a single property but a portfolio, for which the differences are partly smoothed out.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

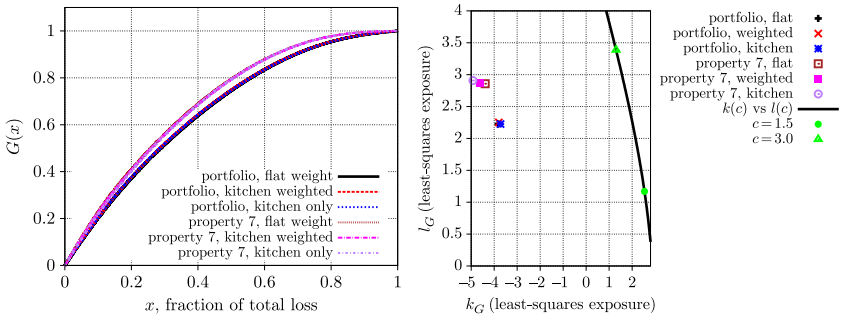


FIGURE C2: (Left) The exposure curve for property 7 and the portfolio for different cases: "flat weight" (all nodes have the same probability of being the origin of the fire; "kitchen only" (all the fires start in the kitchen); "kitchen weighted" (there is a 50% probability that the fire starts because of cooking appliances in the kitchen). Note that all the curves related to the portfolio overlap quite closely, and so do all the curves for Property 7. (Right) The MBBEFD parameters for the various cases depicted on the left.

KS distance	Value = 1, flat weight	Value~U(0,1), flat weight	Value~U(1/2,1), flat weight	Value = 1, kitchen weighted	Value = 1, kitchen only
Value = 1, flat weight		1.3%	0.2%	1.1%	1.3%
Value~U(0,1), flat weight			1.3%	0.3%	0.3%
Value~U(1/2,1), flat weight				1.1%	1.3%
Value = 1, kitchen weighted					0.2%
Value = 1, kitchen only					

RMS distance	Value = 1, flat weight	Value~U(0,1), flat weight	Value~U(1/2,1), flat weight	Value = 1, kitchen weighted	Value = 1, kitchen only
Value = 1, flat weight		0.9%	0.1%	0.8%	0.9%
Value~U(0,1), flat weight			1.0%	0.2%	0.1%
Value~U(1/2,1), flat weight				0.8%	0.9%
Value = 1, kitchen weighted					0.1%
Value = 1, kitchen only					

FIGURE C3: (Left) The KS distance between the empirical exposure curves for the portfolio for five different cases. In the figure, the value of each node can be unitary (Value = 1), a random number between 0 and 1 (Value~U[0,1]), or a random number between ½ and 1 (Value~U[0,1/2]). The probability for a node to be the origin of the fire can be the same for all nodes (flat weight), $50\% + 50\%/n$ probability that the fire starts in the kitchen and $50\%/n$ in any of the other $n-1$ nodes (kitchen weighted), or 100% for the kitchen and 0% for all other nodes (kitchen only). (Right) Same calculations but for the RMS distance.

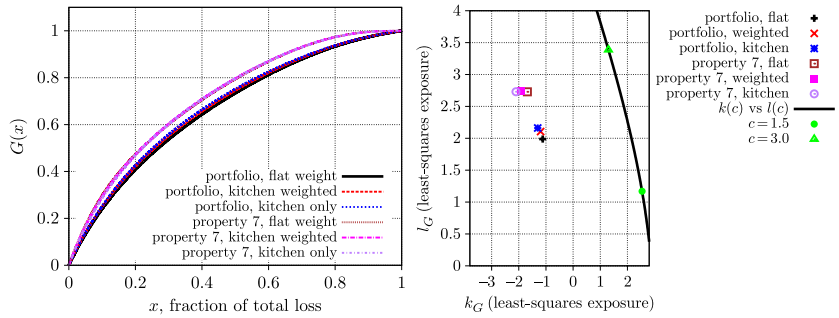


FIGURE C4: (Left) The exposure curve for property 7 and the portfolio for different cases: "flat weight","kitchen only" and "kitchen weighted". As in the case of Figure C2, all the curves related to the portfolio overlap quite closely, and so do all the curves for Property 7. (Right) The MBBEFD parameters for the various cases depicted on the left.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

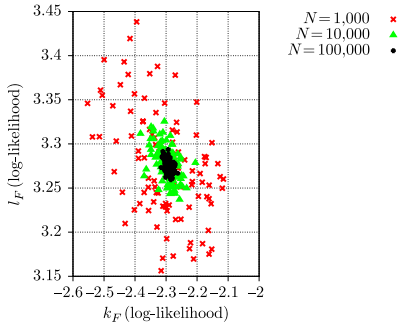


FIGURE D1: Spread of results for the simulation of the parameters k, l as a function of the number of simulated scenarios.

Figure C4 shows the effect of the different treatments of the kitchen for the random time propagation approach (Section 3). The conclusions for this case are very much in line with what stated above for the graph-theoretic approach (Section 2).

APPENDIX D. ACCURACY OF MONTE CARLO SIMULATION

Most of the results in Sections 2 and 3 are based on Monte Carlo simulation. For example, the severity and exposure curves for typical residential structures and a typical residential portfolio were based on simulating 100,000 scenarios (see Figure 10). It is therefore important to estimate the typical simulation error made in the estimation of the parameters. This can be done by repeating the simulation several times and measure the volatility of the parameters calculated in different rounds of simulation.

As an example, Figure D1 shows the spread of the results for the MBBEFD parameters for the case of Property 7 across different simulations for different numbers of simulated scenarios, using the random time fire propagation model of Section 3 (the results for the random graph model of Section 2 are quite similar). The simulation error can then be calculated as the standard deviation of the estimate. For $N = 100,000$, this leads to the following estimates of the parameters k, l:

$$k \approx -2.30 \pm 0.05, \quad l \approx 3.28 \pm 0.02$$

Figure D1 also shows that a 10× increase in the number of simulations roughly leads to a 3× increase in accuracy.

APPENDIX E. ON THE ADEQUACY OF SWISS RE'S *C* CURVES AS A MODEL FOR EXPOSURE CURVES BASED ON RANDOM GRAPHS

As shown in Sections 2.4.3. and 3, the MBBEFD model provides a good fit for the empirical exposure curve based on the random graph model and the random time propagation model

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

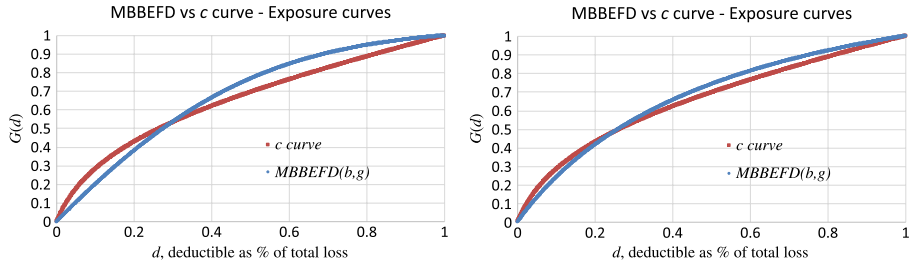


FIGURE E1: (Left) A comparison of the best-fit MBBEFD exposure curve and the best-fit c curve for the portfolio for the random graph simulation method as illustrated in Figure 10. (Right) Same comparison for the random time propagation method as illustrated in Figure 13.

respectively, with KS distance in the order of 1%. However, in both cases the parameters of the MBBEFD model appear to be far from the values associated with Swiss Re's c curves (Guggisberg, 2004).

Figure E1 (left) shows a comparison of the best-fitting MBBEFD exposure curve for the whole portfolio obtained for the random graph case (Section 2.4.3) and illustrated in Figure 10 (parameters: $k = -3.97, l = 2.41$) with the best-fitting c curve ($c = 2.14$, which corresponds to $k = 2.09, l = 2.10$). The two curves are significantly different: the RMS and KS distance are quite large (5.9% and 8.3% respectively). The main difference comes from the fact that while the best-fitting c curve is more concave at the beginning and straightens out towards the end, the best-fitting MBBEFD does exactly the opposite, pointing to the fact that the best-fitting MBBEFD curve has a smaller concentration of smaller losses. The probability of a total loss is not too dissimilar (8.3% for the best-fitting MBBEFD against 10.9% for the best-fitting c curve).

Figure E1 (right) shows the same comparison for the random time case (Section 3), illustrated in Figure 13. The parameters for this case are $k = -1.13, l = 2.00$ for the MBBEFD curve and $c = 2.13$ (which corresponds to $k = 2.10, l = 2.09$) for the c curve. In this case the difference is not as large but the two curves are still not sufficiently close. The qualitative difference in behaviour (larger concavity for the c curve at the beginning) is the same as for the random graphs model, but less pronounced. The RMS and KS distance are 3.4% and 4.7% respectively (smaller but still substantial). The probability of total loss is similar (11.7% for the MBBEFD curve, 10.8% for the c curve).

Overall, the main message for both cases is that Swiss Re's c curves are not a suitable model for our simulated curves based on property graphs (for which the MBBEFD approximation is a good proxy). As mentioned in Section 5.1, there can be several reasons for this. One possibility is that the graph-theoretic approach to building exposure curves is unrealistic. Another possibility is that c curves (derived from old Swiss fire data) are inadequate to represent modern UK houses. The truth might be a combination of these explanations.

A look at the way in which the c curves differ from the best-fitting MBBEFD model provides some early indications of where to look. The fact that c curves appear to have a bigger concentration of small losses is not altogether surprising given that it is not possible to have a loss smaller than the smallest value of the units in the graph. E.g. for our portfolio, where the largest graph has 15 nodes, the smallest possible damage ratio is (in the case where all rooms have the same value) $1/15 \sim 6.7\%$ and the exposure curve is a straight line below that. In order to obtain a behaviour closer to the c curves in this respect (if desired) our model

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

may need to be modified to allow for partial losses for each unit/node (see Section 5.2.1 for some suggestions).

The other observation made above is the fact that our model has a higher concavity on the right-hand side of the curve, implying that there is a higher probability for losses that destroy a large proportion of a building. This is a natural consequence of the fact that connected components that are large but do not invade the whole graph have a substantial probability of occurring with our current choice of parameters. Further empirical research is needed to test whether this effect is observed in practice or whether amendments to the model are needed.

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2019.4 Published online by Cambridge University Press