

Pro-Diluvian: Understanding Scoped-Flooding for Content Discovery in Information-Centric Networking

Liang Wang
University of Cambridge, UK
lw525@cam.ac.uk

Suzan Bayhan
University of Helsinki, Finland
bayhan@hiit.fi

Jörg Ott
Aalto University, Finland
jo@netlab.tkk.fi

Jussi Kangasharju
University of Helsinki, Finland
jakangas@helsinki.fi

Arjuna Sathiaselan
University of Cambridge, UK
as2330@cam.ac.uk

Jon Crowcroft
University of Cambridge, UK
jac22@cam.ac.uk

ABSTRACT

Scoped-flooding is a technique for content discovery in a broad networking context. This paper investigates the effects of scoped-flooding on various topologies in information-centric networking. Using the proposed ring model, we show that flooding can be constrained within a very small neighbourhood to achieve most of the gains which come from areas where the growth rate is relatively low, i.e., the network edge. We also study two flooding strategies and compare their behaviours. Given that caching schemes favour more popular items in competition for cache space, popular items are expected to be stored in diverse parts of the network compared to the less popular items. We propose to exploit the resulting divergence in availability along with the routers' topological properties to fine tune the flooding radius. Our results shed light on designing efficient content discovery mechanism for future information-centric networks.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Network communications; C.4 [Performance of Systems]: Modeling techniques

General Terms

Theory; Design; Performance

Keywords

Information-Centric Networking; Scoped-flooding; Content Discovery; Optimisation; Graph Theory

1. INTRODUCTION

Content, especially popular content, in an information-centric network (ICN) [1–4] may reside “anywhere”, there-

fore the distribution efficiency heavily relies on the effectiveness of content discovery mechanisms. Considering the gap between large content objects and scarce router resources, designing intelligent content discovery to balance protocol simplicity, computational complexity and traffic overhead is crucial in every ICN architecture.

Content discovery is generally achieved by resolution-based [2–8] or routing-based [1, 9, 10] solutions. Resolution-based discovery is a deterministic solution which maps requesters with providers at rendezvous points. The rendezvous point can be either statically configured or referred by proper content addressing [3, 7, 8]. Though resolution-based discovery has relatively small traffic footprint, its performance may degrade quickly in face of large and dynamic content demands. On the other hand, the routing-based discovery usually provides a probabilistic solution. The chances of finding the content can be improved by exploring a larger area of the network, i.e., via collaboration or flooding. In practice, naive network-wide flooding is rarely used due to its significant traffic overhead. A flooding operation is usually constrained within a well-defined neighbourhood (or scope) which is often referred to as *scoped-flooding*. Technically, such constraint on the neighbourhood size is achieved by setting a hop limit for each flooding (e.g., TTL limit).

The use of flooding is based on the following considerations. First, flooding can significantly reduce the protocol complexity and simplify the design, which is very desirable in an unstable environment [11]. Second, in addition to the well-known temporal locality, user requests also possess strong spatial locality [12]. The two localities together indicate that it is highly likely to discover a popular content among nearby neighbours. Third, flooding can reduce the state maintained in the network for a routing-based discovery [13]. Fourth, the communication between close neighbours is relatively cheap (regarding delay, transmission cost and etc.) compared to using backhauls in many cases. Therefore, flooding remains as the default fallback strategy for content discovery if normal forwarding fails in CCNx [14], and also used in various routing and caching designs [10, 13, 15–17, 41].

Despite its wide application (e.g., in ICN [1, 41], P2P [18, 19], MANET [20, 21]), a thorough understanding of how scoped-flooding impacts content discovery is still lacking. More precisely, the following key questions are awaiting answers: (1) what is the optimal radius of scoped-flooding? (2) where do most of the gains come from in a network? (3)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

Accepted to ACM ICN'15, September 30–October 2, 2015, San Francisco, CA, USA.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

how do topological properties of a network impact scoped-flooding? The answers will shed light on designing more intelligent strategies by *flooding for the proper content at the right place with the optimal radius*.

In this paper, to address the aforementioned problems, we propose a node-centric, ring-based model to analyse scoped-flooding. Based on the ring model, we first investigate neighbourhood growth model on general network topologies. The results show that average growth rate increases at least exponentially and can be well estimated using the information within 2-hop neighbourhood. Along with Bayesian techniques, we solve the optimal radius problem and further compare two flooding strategies (static and dynamic) on specific network models.

Specifically, our contributions are:

1. We perform a theoretical analysis on the effects of scoped-flooding using the proposed node-centric ring-based model.
2. The analytical results along with the evaluations show optimal flooding radius is very small (no more than 3 hops).
3. Most of the gains of scoped-flooding are from very small neighbourhoods located at network edges, indicating flooding is more proper at network edge instead of core.

2. SYSTEM MODEL

We assume an information-centric network whose topology is represented with a graph $G = (V, \rho)$, where V is a set of nodes characterized with degree distribution ρ . ρ_k denotes the probability that a node has exactly degree k . The distribution can be arbitrary. For a node v_i , we organize its neighbourhood into r concentric rings according to the lengths of shortest paths between v_i and its neighbours. We denote n_r as average number of r -hop neighbours on the r^{th} ring. We refer to this model as node-centric ring-based model, or simply a *ring model*. In reality, nodes may break down resulting in lost messages. We model the stability with $\gamma \in (0, 1]$ which denotes the probability that a router is up and working properly, i.e., the reliability rate. Equivalently, $(1 - \gamma)$ denotes the failure rate.

Designing a fully-fledged protocol is out of the scope of this paper. Instead, we briefly describe general flooding behaviours in the following. Nodes in a network receive requests from either directly connected clients or neighbours. We exclude clients from the model and focus only on core network. Whenever a request arrives, a node first looks for a match in its local cache. If the node cannot find the requested content locally, it decides whether to initiate a scoped-flooding before simply forwarding the request to the next hop along the path to original content providers. The flooding is constrained within a r -hop neighbourhood by maintaining a hop counter in packet header. The node terminates the flooding if the hop counter reaches r . If the content is discovered, we assume the content can always return to the initial flooding node in a reverse route similar to CCNx. To prevent loops, nodes do not re-flood the messages they have seen before.

For simplicity, we do not consider the case of partially matched content. A node i either has the exact requested content or not. The value of response R is described as an

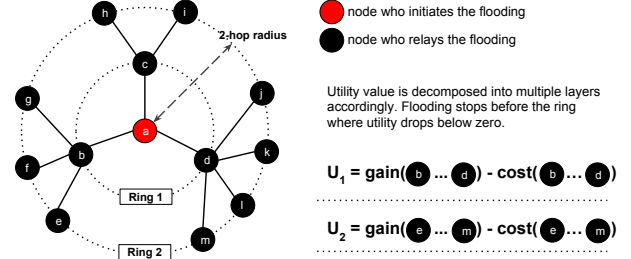


Figure 1: In a ring model, to estimate whether the utility on the $(r + 1)^{th}$ ring drops below zero, a node on the r^{th} ring needs to know three pieces of information (1) which ring it is; (2) how its neighbourhood grows; (3) content availability.

i.i.d random variable which follows a Bernoulli distribution $R \sim \text{Bernoulli}(p)$, with 1 representing a successful discovery and 0 otherwise. p is often referred to as *content availability*. Note that p itself can follow different distributions to model content availability (e.g., Zipf or Weibull). By definition, we let $q \triangleq 1 - p$ denote the probability of failing to find the matched content on a node. Fig. 1 illustrates the ring model under our investigation. Note that there will be less symmetry at network edges but the model remains the same.

There are many resource constraints in a network such as energy, bandwidth and storage. In our model, we use c to represent the cost induced by receiving and processing flooding requests. Besides, consecutive requests may also impact content delivery in terms of added queueing and processing delay which we assume to be roughly proportional to the number of messages. However, the cost can increase faster and requests may be dropped in a busy network. To generalise, the cost is modelled as a linear function of number of nodes involved in a flooding.

3. NEIGHBOURHOOD GROWTH MODEL

The first step to solve optimal radius is to understand how neighbourhood grows as a function of flooding radius. Newman derived this functional relation in [22] using a general graph model $G = (V, \rho)$. We recap briefly the major steps of the derivation in Section 3.1, based on which we investigate two specific types of networks, i.e., random networks and scale-free networks. Then we examine the accuracy of estimates on both synthetic and realistic networks.

3.1 Average Number of r -Hop Neighbours

The effective topology due to a flooding can be viewed as a distribution tree. On non-trivial topologies, such a tree cannot be easily decomposed into multiple linear models (from root to leaves). We apply ring model to organize the neighbourhood of node v into r concentric rings according to the neighbour's distance to v so that we can study the growth ring by ring. Calculating n_1 , namely the average number of directly connected neighbours, is trivial. Let $\langle k \rangle$ denote the mean of a given degree variable k . Average number of 1-hop neighbours equals the node's average degree as follows:

$$n_1 = \langle k \rangle = \sum_{k=0}^{\infty} k \rho_k. \quad (1)$$

However, calculating n_r ($r \geq 2$) is not as straightforward as n_1 since the degree distribution of a node's neighbour is

not the same as the general degree distribution of the whole network [24]. Let v_j be one of v_i 's next-hop neighbours and τ_k be the probability of v_j having k emerging edges which lead to k new next-hop neighbours. Note that we exclude the edge leading back to v_i from v_j since it does not contribute to new nodes. The results in [25] show τ_k is proportional to both v_i 's degree and general degree distribution of the network. The reason is the edges of a high-degree node have a higher chance to connect to any given edge in the network. The probability of v_j having k new next-hop neighbours is:

$$\tau_k = \mathbf{Pr}[deg(v_j) = k|\rho] = \frac{(k+1)\rho_{k+1}}{\sum_m m\rho_m}.$$

Therefore, the average number of new nodes from v_j is:

$$\begin{aligned} \sum_{k=0}^{\infty} k\tau_k &= \frac{\sum_{k=0}^{\infty} k(k+1)\rho_{k+1}}{\sum_m m\rho_m} = \frac{\sum_{k=0}^{\infty} k(k-1)\rho_k}{\sum_m m\rho_m} \\ &= \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}. \end{aligned}$$

Because we did not assume v_i is on any specific concentric ring except $r \geq 2$, we can use the same τ_k and the same logic above to calculate arbitrary r -hop neighbours. Namely, n_r equals the average number of nodes on the $(r-1)^{th}$ ring multiplied by their average out-degree to the r^{th} ring.

$$\begin{aligned} n_r &= n_{r-1} \sum_{k=0}^{\infty} k\tau_k = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} n_{r-1} \\ &= \left[\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right]^{r-1} \cdot \langle k \rangle \end{aligned} \quad (2)$$

Eq.(2) shows that the number of r -hop neighbours is a function of the degree variable. Using eq. (2), we can calculate $n_2 = \langle k^2 \rangle - \langle k \rangle$. As we know $n_1 = \langle k \rangle$, by applying the replacement recursively, we can rewrite eq. (2) as below, which eventually leads us to the same function found in [22].

$$n_r = \left[\frac{n_2}{n_1} \right]^{r-1} \cdot n_1. \quad (3)$$

Eq.(3) shows that n_r can also be expressed as a function of the ratio between average number of 2-hop and 1-hop neighbours. The neighbourhood size only converges if there are fewer 2-hop neighbours than 1-hop ones, i.e., $\frac{n_2}{n_1} < 1$, which actually implies the network has multiple components with high probability. We define *neighbourhood growth rate* β as

$$\beta \triangleq \frac{n_2}{n_1} \triangleq \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}. \quad (4)$$

Note that the derivation above applies to any general network of arbitrary degree distributions. The result gives an interesting implication on the topological inference, which says a node can approximate β by utilizing the local knowledge within its 2-hop neighbourhood. In the following, we focus on the growth rate β and study two specific network models: random network and scale-free network. Both are prominent in networking research due to their representativeness of many realistic networks [27–29]. Specifically, empirical evidence shows mobile and opportunistic networks can be either random [29] or scale-free [30], whereas fixed and wired networks are mostly scale-free [27, 28].

3.2 Case 1: Random Networks

Random networks have a binomial degree distribution $B(|V|, \rho)$ which is given by the following formula [23]

$$\rho_k = \binom{|V|-1}{k} \rho^k (1-\rho)^{|V|-k-1}.$$

For very big $|V|$ and small ρ , the binomial distribution above converges to the Poisson distribution in its limit. Then, the degree distribution ρ_k becomes:

$$\lim_{|V| \rightarrow \infty} \rho_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}.$$

For calculating β in eq.(4), we need to derive the second moment of random variable k , i.e., $\langle k^2 \rangle$. Using Touchard polynomials¹, the r^{th} moment of a variable with Poisson distribution can be calculated as eq. (5) shows. $\{ \left\{ \begin{smallmatrix} r \\ k \end{smallmatrix} \right\} \}$ denotes *Stirling numbers of the second kind* [23] which represents the number of ways to partition a set of r objects into k non-empty subsets, and is known for calculating $\langle k^r \rangle$.

$$\langle k^r \rangle = e^{-\langle k \rangle} \sum_{k=0}^{\infty} \frac{\langle k \rangle^k \cdot k^r}{k!} = \sum_{k=1}^r \left\{ \begin{smallmatrix} r \\ k \end{smallmatrix} \right\} \langle k \rangle^k \quad (5)$$

Combining eq. (5) and eq. (2) yields

$$n_2 = \left\{ \begin{smallmatrix} 2 \\ 2 \end{smallmatrix} \right\} \langle k \rangle^2 + \left\{ \begin{smallmatrix} 2 \\ 1 \end{smallmatrix} \right\} \langle k \rangle - \langle k \rangle = \langle k \rangle^2. \quad (6)$$

Similarly, by applying the replacement recursively, we get

$$n_r = \langle k \rangle^r \implies \beta = \langle k \rangle \quad (7)$$

Eq. (7) shows that $n_1, n_2, n_3 \dots$ form a geometric series. The growth rate is $\beta = \langle k \rangle$. It is worth noting that many topological properties (e.g., average degree, density etc.) are homogeneous on random networks. In other words, a randomly chosen sub-network possesses similar characteristics as the whole network which is also known as self-similarity.

3.3 Case 2: Scale-free Networks

Although random networks give a very neat form of growth rate, many realistic networks are scale-free and the node degree follows a power-law distribution, i.e., $\rho \propto k^{-\alpha}$ with $\alpha > 2$ [27, 28, 30]. For a power-law distribution, the r^{th} moment of random variable k equals:

$$\langle k^r \rangle = k_{min}^r \cdot \frac{\alpha - 1}{\alpha - 1 - r} \quad \forall \alpha > r + 1 \quad (8)$$

Note that a power-law distribution is extremely right-skewed and has a heavy tail. Only the first $\lfloor \alpha - 1 \rfloor$ moments exist, the other moments are infinite. If we plug eq.(8) into eq.(2) and let $k_{min} = 1$, the growth rate equals:

$$\beta = \frac{1}{\alpha - 3} \quad \forall \alpha > 3 \quad (9)$$

Eq. (9) means that though most real-life networks have a well-defined average node degree, their variance is infinite, which further indicates the growth rate β is unbounded.

¹We can also use moment generating functions for a Poisson random variable with parameter λ , i.e., $M_X(t) = e^{\lambda(e^t - 1)}$, and we derive $\langle k^2 \rangle$ by calculating $M_X''(t=0)$. This gives us: $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$. $\langle k^r \rangle$ can be calculated using higher order moments similarly.

Table 1: Overestimation of the model at each hop for various network graphs. V : Number of nodes and E : Number of nodes in the generated instance of the graph, l : average path length. Shaded cells represent the cases where the error is below 0.20.

Id	Topology	V	E	$\langle k \rangle$	l	Clustering	Overestimation of the model				
							$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$
1	Random	339	338	1.994	23.07	0	0.327	1.046	2.359	4.692	9.092
2	Random	8030	9761	2.431	12.03	0	0.152	0.371	0.642	0.972	1.399
3	Random	9426	15068	3.197	8.30	0.00040	0.060	0.130	0.212	0.332	0.565
4	Random	9811	20073	4.091	6.75	0.00049	0.023	0.053	0.106	0.259	0.873
5	Random	9928	25060	5.048	5.88	0.00048	0.004	0.017	0.079	0.419	2.79
6	Random	9989	35020	7.011	4.95	0.00066	0.003	0.030	0.229	2.139	54.124
7	Scale-free, $\alpha = 3.24$	7141	9648	2.70	7.88	0.00057	0.093	0.271	0.529	1.069	2.599
8	Scale-free, $\alpha = 3.35$	5869	7347	2.50	8.66	0.00076	-0.115	-0.174	-0.194	-0.16	0.013
9	Scale-free, $\alpha = 3.50$	5960	7357	2.47	8.99	0.00013	-0.356	-0.555	-0.68	-0.757	-0.794

For $3 < \alpha < 4$, the growth rate is bounded but the neighbourhood size never converges. It is also interesting to notice when $\alpha > 4$, n_r converges to zero at its limit $r \rightarrow \infty$. The reason is the existence of super hubs with extremely high degrees which strengthens the small-world effect and makes the network diameter extremely short. We refer to [23] for more thorough and interesting discussions on graph topological properties. For both random network and scale-free network, we can see neighbourhood growth is at least exponential which sheds light on the flooding strategy design.

3.4 Accuracy on Estimating β

To assess the model accuracy, we generate random and scale-free topologies for which we calculate the actual average neighbourhood at each hop distance, i.e., \bar{n}_r . To derive the n_r estimated by the model, we first find the parameter of a corresponding degree distribution, i.e., Poisson for Erdős-Rényi random graph and power-law for scale-free graph, by maximum likelihood estimation.² After finding the distribution parameter, we calculate n_r using eq.(7) or eq.(9) and compute the deviation from \bar{n}_r by $(n_r - \bar{n}_r)/\bar{n}_r$. For both topologies, we set the number of nodes to $N = 10000$. If a generated network is not connected, we use the largest component hence V can be smaller than N . The link probability parameter ρ determines the number of edges in an Erdős-Rényi graph, similar to the exponent α in a scale-free network.

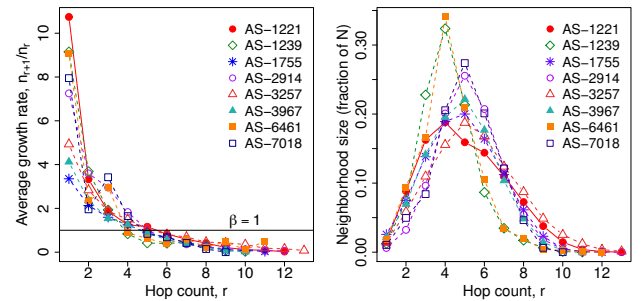
Table 1 summarizes the network properties along with the deviation, i.e., overestimation ratio. $r = 1$ is excluded as it converges to 0 for all settings. For almost every setting, the model overestimates the reality only slightly for $r = 2$ and $r = 3$. For $V = 339$, we attribute the deviation to both the finite size effect as well as the absence of random graph property, i.e., the network does not exhibit Poisson degree distribution as the model assumes. Increasing hop count makes the model deviate significantly from the reality, especially when $r \geq l$, which is expected as a result of finite size of the networks. For $r = 4$, the model captures the reality quite well for large V and moderate $\langle k \rangle$ – the region where the random graph property exists but the network is not so densely connected. The deviation is higher for the settings with higher $\langle k \rangle$ due to higher clustering and smaller network diameter.

For scale-free networks, eq.(9) may either underestimate or overestimate depending on the power-law exponent α .

²For scale-free networks, we use the method described in [31].

For $\alpha \approx 3$, the expected growth rate is very large resulting in overestimation in neighbourhood (e.g., topology-7 in Table 1). For $\alpha > 3$, the estimated growth is more stable which leads to underestimation of the real growth, e.g., topology-8 and topology-9. We attribute this dispersion to the diversity of the degree distribution in a scale-free network and limitations of our model to represent this diversity accurately.

The ISP networks are smaller, ranging from a couple of hundreds to thousands of nodes [27], which results in a slower growth after certain hops. To understand this effect, we derive the growth rate at r^{th} hop as $\beta_r = \frac{n_{r+1}}{n_r}$ and plot them in Fig.2 for eight ISP networks. Recall that in the analysis we have a single β value for the whole networks with $N \rightarrow \infty$. As the figure shows, the growth rate decreases with increasing hop due to the finite size of the network. Although the growth rate is a decreasing function of r , we can observe in Fig. 2 that the neighbourhood keeps growing for several hops, e.g., $r \approx 5$. β_r takes values below 1 for r greater than average path length that varies between 3.36 hops to 5.51 hops. In general, the neighbourhood growth model performs very well within a moderate scope on both synthetic and realistic networks.



(a) Growth rate at each hop. (b) Neighbourhood size growth.

Figure 2: Change in neighborhood in real ISP networks.

4. OPTIMAL FLOODING RADIUS

Based on the previous growth model, we continue our study on calculating the optimal flooding radius in two cases: with and without prior knowledge on content availability.

4.1 Effective Nodes

Since nodes may be up or down, we let γ denote the probability that a node is up, namely a node's reliability rate. We define the *effective nodes* \hat{n}_r as the nodes that are working and also reachable on the r^{th} ring. Since only the effective nodes contribute to flooding messages, i.e., improving content discovery, it is crucial to know the growth of effective nodes for a specific γ in order to derive the optimal radius.

Given a node has n_1 1-hop neighbours, its effective 1-hop neighbours equals $\hat{n}_1 = \gamma n_1$ by assuming a node's state (up or down) is independent of each other. Given growth rate β , the effective 2-hop neighbours equals $\hat{n}_2 = \beta \gamma^2 n_1$. Similarly, we can calculate \hat{n}_3 using \hat{n}_2 . Applying iteratively, we calculate the effective nodes on the r^{th} ring as follows:

$$\hat{n}_r = (\beta \gamma)^{r-1} \gamma n_1 = \gamma^r n_r. \quad (10)$$

It is easy to see the similarity between eq.(3) and eq.(10). In fact, $\hat{n}_1 = \gamma n_1$ is the effective 1-hop neighbours and $\beta \gamma$ can be viewed as effective growth rate given nodes may fail with certain probability $(1 - \gamma)$. For low reliability rates, the gap between the number of effective nodes and the r -hop neighbourhood will quickly increase with an increasing r .

4.2 Content Availability as A Priori

The purpose of flooding is to increase the chance of discovery by visiting enough nodes. Given n visited nodes, the probability of finding the content of availability p equals $(1 - q^n)$ which we use to represent the gain from a flooding. On the other hand, a bigger n also introduces larger cost which limits the utility U as eq.(11) shows:

$$U = (1 - q^n) - n \cdot c. \quad (11)$$

$-U$ in eq.(11) is apparently convex as an exponential function is convex and the linear combination of convex functions preserves convexity. To maximise U , the optimal number of nodes n^* we need to visit can be calculated as below:

$$U'(n) = 0 \implies -q^n \cdot \ln q - c = 0 \implies n^* = \frac{\ln c - \ln \ln q^{-1}}{\ln q}.$$

n^* represents the optimal total number of nodes. Using eq.(10), we can calculate the optimal radius by summing up the effective nodes from ring 1 to r then solving the equation below.

$$\sum_r \hat{n}_r = \sum_r (\beta \gamma)^{r-1} \gamma n_1 = n^*$$

4.3 Inferring the Content Availability

We previously assumed that the content availability p is known *a priori*. Technically, we can set up monitoring nodes to sample request streams. However, monitoring can be expensive and sometimes may not even be feasible. Nevertheless, the probability of finding a specific content in a neighbourhood is a good indicator for its actual availability, since the more popular a content is, the more probable it is to find it among nearby neighbours. We use the Bayesian technique proposed in [32] to estimate content availability. Eq.(12) is the probability density function of p conditioned on previous i negative (i.e., unsuccessful) queries.

$$f(p|i) = \frac{\Pr(i|p) \cdot f(p)}{\int_0^1 \Pr(i|p) \cdot f(p) dp} \quad (12)$$

Because $\Pr(i|p) = q^i$, if we use the Bernoulli distribution and let $f(p) = 1$, then we have

$$f(p|i) = \frac{q^i}{\int_0^1 q^i dp} = (i+1)q^i.$$

After getting the posterior of p , we can calculate the expected p after i negative queries as below

$$\langle p \rangle = \int_0^1 p(i+1)q^i dp = \int_0^1 (i+1)(1-q)q^i dq = \frac{1}{i+2}. \quad (13)$$

Note that neither p nor q appears in eq.(13). The derivation above gives a very clean estimation of content availability especially when monitoring is not possible or the content has never been observed before.

4.4 Content Availability as Posteriori

Without prior knowledge on content availability, we cannot apply the conventional optimization as that in Section 4.2. Even with the Bayesian inference introduced in Section 4.3, deciding the optimal radius can be difficult, especially when the request comes from directly connected clients or does not carry any information about the number of nodes it has traversed. To get around this challenge, we let a node flood its 1-hop neighbours by default to bootstrap the inference on p . Then we consider the utility of each ring separately and adaptively adjust the estimate of p on every ring. The general mechanism can be summarized as follows:

1. If a request does not contain useful information for estimating the availability (e.g., number of nodes queried), a node initiates a flooding to its directly connected neighbours. A flood message carries 3 pieces of information: the node's local growth rate $\beta = \frac{n_2}{n_1}$; number of 1-hop neighbours n_1 ; a counter r to record the number of hops it has travelled.³

2. When a node receives a flood message, it first estimates the availability p using β , n_1 and r embedded in the message by assuming the requested content cannot be found so far (within r -hop neighbours). More particularly, as follows:

$$\langle p \rangle = \frac{1}{\beta^{r-1} \gamma^r \cdot n_1}$$

Using this estimated p , the node then estimates the potential utility of the next ring. Based on the estimated utility, the node decides whether to continue the flooding or terminate.

More specifically, the overall utility of scoped-flooding is decomposed according to our ring model. Given that R_r and C_r represent the aggregated gain and the aggregated cost on the r^{th} ring respectively, the net utility of a flooding is:

$$U = \sum_r U_r = \sum_r (R_r - C_r).$$

According to eq.(10), the average cost on the r^{th} ring is:

$$E(C_r) = \hat{n}_r c = \gamma^r n_r c$$

³Note that β , n_1 , and n_2 here refer to the **local** properties of a specific node instead of the global average. We avoid new notations because the following derivation on optimal radius applies to both local and global cases which is independent on the parameters plugged in. As we will show in Section 5, Dynamic flooding uses local parameters while Static uses global ones.

and the average value of gross gain R_r is:

$$E(R_r) = 1 \cdot (1 - q^{\gamma^r n_r}) + 0 \cdot q^{\gamma^r n_r} = 1 - q^{\gamma^r n_r}.$$

The net utility value from the r^{th} ring therefore can be expressed as the difference between $E(R_r)$ and $E(C_r)$, namely

$$U_r = E(R_r) - E(C_r) = (1 - q^{\gamma^r n_r}) - \gamma^r n_r c. \quad (14)$$

An intermediate node forwards the flooding message to its next-hop neighbours only if the next ring can bring positive net utility, which can be easily tested with eq.(14). The flooding radius should stop increasing whenever the expected utility of next ring falls below zero. Technically, this is solved by calculating the root of eq.(14) which is the maximum number of effective nodes on the r^{th} ring. Note that $U_r \geq 0$ indicates $c \leq \frac{1 - q^{\gamma^r n_r}}{\gamma^r n_r}$ which provides a clear decision boundary on whether to continue a flooding operation.

Given $\gamma = 1$, which indicates a stable network of no failures, the root of eq.(14) above reduces to:

$$1 - q^{n_r} = n_r c.$$

The mixture of exponential and polynomial functions can be solved with the Lambert W function, which gives:

$$n_r^* = -\frac{1}{\ln q} W_k\left(\frac{\ln q}{c} e^{\frac{\ln q}{c}}\right) + c^{-1}. \quad (15)$$

n_r^* represents the maximum number of nodes that the r^{th} ring can have in order to keep the cost smaller than the gain. By plugging eq.(15) into eq.(2), we can easily derive the optimal flooding radius r^* as a function of cost, content availability and neighbourhood growth rate.

$$n_r = \beta^{r-1} n_1 = n_r^* \implies (r-1) \ln \beta + \ln \langle k \rangle = \ln n_r^* \quad (16)$$

$$\implies r^* = \frac{\ln n_r^* + \ln \beta - \ln \langle k \rangle}{\ln \beta} \quad (17)$$

Given $0 < \gamma < 1$, we have the same derivation except \hat{n}_r replaces n_r in eq.(16). After some manipulations, we have

$$\hat{n}_r = \gamma^r \beta^{r-1} n_1 = n_r^* \implies r^* = \frac{\ln n_r^* + \ln \beta - \ln \langle k \rangle}{\ln \gamma + \ln \beta}. \quad (18)$$

Obviously, $\gamma = 1$ indicates $\ln \gamma = 0$, then eq.(18) reduces to eq.(17) as expected. Since $\ln \gamma$ is a monotonically increasing function and only appears in the denominator of eq.(18), r^* is hence a decreasing function of γ . In practice, eq.(18) means the flooding radius tends to be bigger in an unstable network to achieve the same gain. On the other hand, for a given reliability rate γ , the optimal radius r^* is a decreasing function of the growth rate β .

5. TWO FLOODING STRATEGIES

We first discuss the design rationale behind a scoped-flooding, then introduce two strategies for the later comparison.

5.1 Design Guidelines

A good flooding strategy requires that: (1) a node is aware of its neighbourhood with an accurate topological inference; (2) a node is aware of content availability with an accurate statistical inference on user request streams. These two awareness (solved in Section 3 and 4 respectively) together enable a node to decide its optimal flooding radius based on the estimated utility. In addition, the flooding radius should

be adjusted adaptively in different areas according to local topological properties because the network structure may not be homogeneous, i.e., some parts are denser and some parts are sparser (regarding degree distribution). Hence a predetermined radius may lead to suboptimal performance.

5.2 Static Flooding

Static flooding uses a predetermined and fixed flooding radius for all the nodes. The flooding radius is optimized over the whole network topology by e.g., the network operator for each availability value. The average growth rate is calculated using the average number of 1-hop neighbours and 2-hop neighbours of the whole network then plugged into either eq.(17) or eq.(18) to derive the optimal radius. Therefore static flooding ignores the heterogeneity of the topological properties in different areas of the network. Static flooding is simple and popular but it is only suitable for random networks wherein the network structure is homogeneous and nodes have similar growth rates. We include static flooding in our evaluation as a baseline for comparison.

5.3 Dynamic Flooding

Compared to static flooding, dynamic flooding is more attendant to the differences among the nodes and it assigns a specific radius for each node individually. Considering that the degree distribution in a scale-free network is not homogeneous, dynamic flooding lets each node use its own 1-hop and 2-hop neighbours to calculate the local growth rate. Then each node optimises locally within its neighbourhood, hence each has its own optimal flooding radius. Such local optimisation strategy takes a node's position in a network into account. Nodes in denser areas tend to have smaller radius while nodes in sparser areas tend to have bigger radius.

For less available content, a node may prefer routing towards the original content provider rather than initiating flooding. By letting $r = 1$, eq.(14) calculates the availability threshold of whether initiating a flooding as below

$$U_1 > 0 \implies q^{\gamma n_1} < 1 - \gamma n_1 c \implies p > 1 - \gamma n_1 \sqrt[1-\gamma]{1 - \gamma n_1 c}.$$

If the availability falls far below the threshold, a node will not flood the request. If content availability is unknown, dynamic strategy floods its 1-hop neighbours by default to bootstrap the inference as described in Section 4.4. As we can expect, without content availability information, dynamic flooding is supposed to introduce more overhead due to its aggressive 1-hop flooding. However, the evaluation in Section 6 shows that such overhead is almost negligible.

6. EVALUATION

We evaluate the two flooding strategies on various topologies to gain a comprehensive understanding of their pros and cons. Our evaluations focus on two network models: random networks and scale-free networks. Both models have a network of 10,000 nodes and 60,000 edges but their degree distributions are different, namely one is Poisson and the other is power-law. We experimented with a large number of network parameters and various availability and cost values to guarantee the robustness and consistency of our claims.

6.1 Impact of Availability and Cost

Fig. 3 depicts the model behaviours with different cost and availability parameters. In Fig. 3a and 3b, the curves

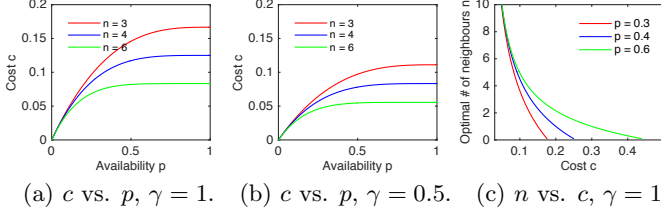


Figure 3: Ring model behaviours with different p and c .

are the decision boundaries below which a node will initiate a flooding for given cost and availability values. The lower cost for higher n_r shows that a node with a large neighbourhood is more parsimonious in flooding compared to a node with a smaller neighbourhood, and only initiates flooding for lower costs. Fig. 3b is a condensed version of Fig. 3a due to setting $\gamma = 0.5$, indicating that an unstable network cannot tolerate high cost values. For both figures, the steep increase in the interval $[0, 0.4]$ indicates the strong preference on high available content in scoped-flooding. Fig. 3c shows when the cost slightly increases from its minimum (i.e., zero), the optimal number of neighbours drops drastically regardless of content availability. After a certain point, e.g., 5 or 6 nodes, the figure shows a slower decrease for content with high availability. Fig. 3c indicates that for higher availability content, it is worth flooding to more neighbours since the content will be discovered with high probability therefore the gain is guaranteed. Whereas for low availability content, even with relatively low cost, the flooding is rather conservative.

6.2 Flooding Radius Distribution

Fig. 4a shows the CDF of nodes' optimal radii using different p on both random (upper figure) and scale-free (lower figure) networks. By increasing p from 0.1 to 0.9, the CDF curves shift towards right indicating high available content is worth large radius. In random network, the shapes of the curves are mostly identical, whereas in scale-free network, the curves are more stretched for higher p values, which indicates more heterogeneity in scale-free networks. Fig. 4b and 4c specifically plot the radius distributions for $p = 0.8$. The scale-free network has smaller flooding radii than the random network in both dynamic and static flooding. The two red vertical lines in Fig. 4c represent the optimal radii of static flooding, i.e., 2.320 for scale-free and 2.783 for random network. For dynamic flooding, the mean and variance of the radii are 2.447 and 0.094 on the scale-free (red area), and 2.805 and 0.014 on the random (blue area). Static flooding ignores the difference of topological characteristics between two nodes. Fig. 4b shows that over 60% of the nodes in scale-free network have a radius less than 2.5 whereas almost all the nodes' radii in random network are bigger than 2.5. In both Fig. 4b and 4c, the left tail of scale-free network is heavier than that of random network due to the existence of high-degree nodes. The radius distribution of random network is more condensed in a smaller range (reflected as a small variance 0.014) because of its homogeneous structure.

There is a relatively strong negative correlation between optimal radius and node's degree as well as optimal radius and node's betweenness centrality. We report the results for betweenness centrality in Fig. 4d. Fig. 4d plots the between-

ness centrality as a function of radius for the scale-free network. Note the logarithmic scale in the axis. We attribute the negative correlation to the nodes with high betweenness centrality that are located in the well-connected parts of the network wherein the link density is very high and therefore the radius is small due to the high growth rate.

6.3 Utility and Its Improvement Distribution

Inspired by [42], fig.5a and 5b plot betweenness centrality as a function of a node's utility. Fig.5b is log-log plot. We observe a strong negative correlation between the two variables. The corresponding Pearson correlation can reach -0.93 and -0.80 for random and scale-free network, respectively. The reason for the negative correlation is that, in the dense area where a node has a high betweenness centrality value, its neighbourhood size is usually big. Although the radius is also small, the node may still include more neighbours than necessary (the optimum) which renders a higher cost and drags down the net utility. Sometimes, even 1-hop neighbours include too many nodes. On the other hand, the growth rate in the sparser area is much lower, so nodes have a better control over the neighbourhood size by fine-tuning their radius leading to smaller cost and better net utility.

To compare dynamic flooding against static one, we let U_{dy} denote the optimal utility achieved by dynamic flooding and U_{st} by static flooding. Then, we calculate the utility improvement as: $\frac{U_{dy} - U_{st}}{U_{st}}$. Fig. 5c plots the CDF of the utility improvement. We notice that dynamic flooding is less effective on random networks, only 10% of the nodes actually improve their performance and over half have less than 10% improvement. Such lower effectiveness of dynamic flooding is due to the homogeneous structure of the random network. As we showed in the previous section, the static optimal radius deviates from the dynamic optimal radius only slightly for a random network. Hence, the improvement in utility is marginal. On the other hand, nodes in scale-free network have much more significant utility improvement, namely about 30% of the nodes are improved, among which over 60% have larger than 10% improvement.

Specifically, we take a closer look at those nodes with improved utility, i.e., the 10% in random and 30% in scale-free network. Fig. 5d and 5e plot local growth rate β as a function of improvement. Note the difference in both X-range and Y-range of the two figures. As for X-range, the utility shows a wider range of improvement in scale-free networks due to the diverse growth rate of the nodes shown on the Y-axis. Scale-free network has a larger β due to hub nodes compared to the random network with more homogeneous node characteristics. Fig.5d shows that the correlation between β and the utility improvement on random network is close to zero, more precisely -0.0031 , indicating that the significance of improvement is irrelevant of a node's growth rate and its position in the network. Meanwhile, such correlation on scale-free network is much stronger, with Pearson correlation being -0.5273 . The results indicate that nodes with high growth rate β are less likely to have significant improvement by using dynamic flooding. The reason is that the optimal radii of the nodes with high β values in both static and dynamic flooding are small and close to each other. However, dynamic flooding usually significantly increases the radius of the nodes with low β .

6.4 Flooding in the Wild

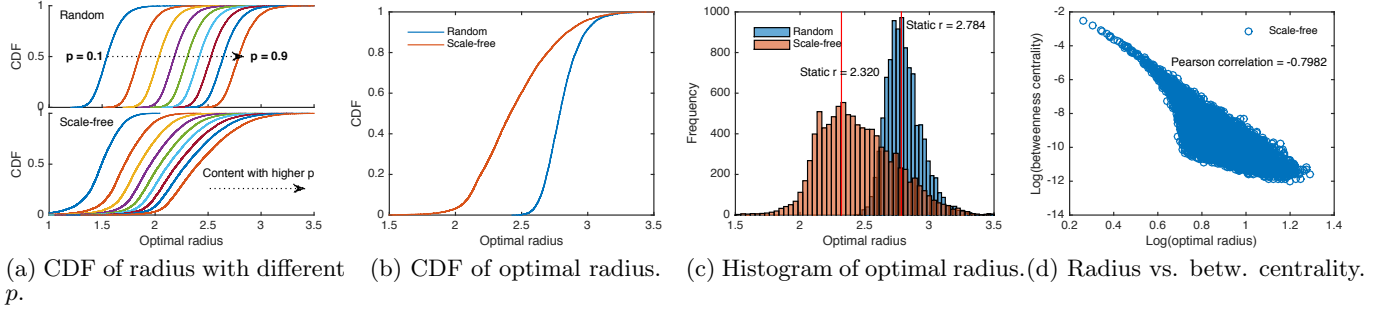


Figure 4: Optimal radius distribution in dynamic flooding, the radius negatively correlates to nodes' betweenness centrality.

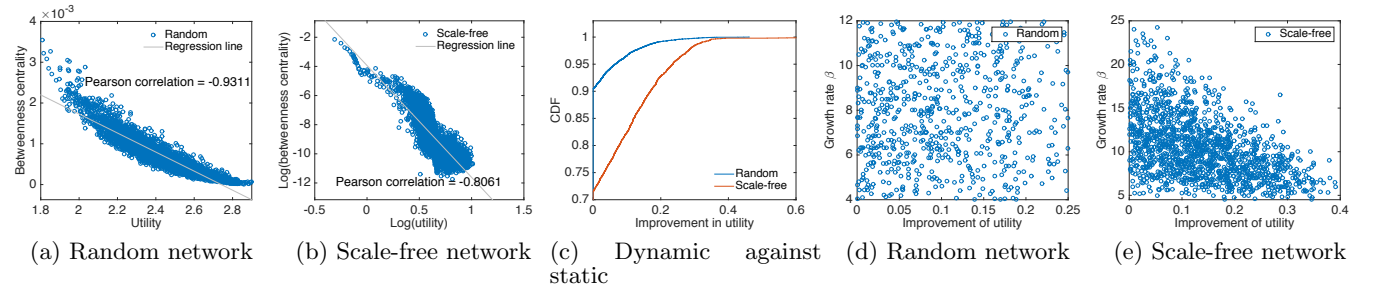


Figure 5: In dynamic flooding, the utility distribution strongly correlates to nodes' positions. However, regarding improvements, the strength of correlation between the significance of improvement and the position depends on the network model.

Table 2: Results for each metric are in the format of *network-wide|static|dynamic* flooding. The cost is measured by the # of flood messages and normalized with the maximum value.

AS	Byte hit rate			Cost			Avg. hops		
	nw	st	dy	nw	st	dy	nw	st	dy
1239	0.44	0.40	0.43	1.0	0.27	0.28	1.90	1.60	1.62
2914	0.49	0.42	0.47	1.0	0.31	0.32	1.75	1.55	1.58
3356	0.42	0.39	0.42	1.0	0.25	0.27	2.02	1.69	1.74
7018	0.47	0.41	0.45	1.0	0.26	0.28	1.87	1.54	1.63
Guifi	0.51	0.44	0.49	1.0	0.22	0.23	1.71	1.32	1.38

To confirm our analysis on realistic networks, we choose four realistic ISPs along with one community network (i.e., Guifi Catalunya region) [27, 28] to compare network-wide, static, and dynamic flooding. Note that dynamic flooding is not aware of content availability but use the inference technique in the evaluations. In case there are multiple components in a network, we use the biggest one. The smallest network (AS 3356) has 3,107 nodes and 6,097 edges while the biggest one (AS 7018) has 9,732 nodes and 10,047 edges. Each node is equipped with a 4 GB cache using LRU for cache replacement, and assigned a pair of geographical coordinates according to the topology traces. The content set is based on the Youtube Entertainment Category trace [33] which contains 1,687,506 objects (average size is 8.0 MB and aggregated size is 12.87 TB). The trace contains video id, length, views, rating and etc. All the nodes of degree 1 are considered as content requesters while 10 to 20 content providers (proportional to the network size) are randomly selected among the nodes in the network. A node cannot be

a requester and a provider at the same time. To take into account both temporal and spatial locality, we use Hawkes process-based algorithm [34] with different *spatial locality factors* $[0, 1]$ to generate user request streams. Locality factor 0 indicates the request pattern reduces to *Independent Reference Model*, 1 indicates high spatial localisation. The “warm-up” period for pre-filling the caches is excluded from an evaluation and the result is averaged over at least 50 runs.

Table 2 reports the results with spatial locality factor 0.5. Although network-wide flooding always achieves the best byte hit rate, the improvement is rather marginal over dynamic flooding (less than 5%). On the other hand, such a small gain in cache hits is at the price of 2 ~ 3 times increase in the number of control messages as the second column shows. Intuitively, without prior knowledge on content availability, the performance of dynamic flooding is mostly affected by network topology and should be worse than static flooding which can exploit such knowledge. The results however show that dynamic flooding consistently outperforms static one, which further attests the effectiveness of Bayesian inference and justifies the design of dynamic flooding. Compared to static one, dynamic flooding has slightly higher cost because it tends to explore more nodes (recall the default flooding to 1-hop neighbours in dynamic flooding), which also explains its gain in byte hit rate. The third column shows the average hops between a requester and the first discovered content. Network-wide flooding has the worst values and static flooding is slightly higher than dynamic one. In all cases, most content are discovered within 2 hops.

We investigate the effects of spatial locality by varying the factor between 0 and 1. For network-wide flooding, spatial locality does not appear to have any noticeable impacts

on the byte hit rate and the cost except that high factor values lead to shorter average hops. On the other hand, higher spatial locality factor improves byte hit rate and average hop count in both static and dynamic flooding. For dynamic flooding, by increasing the locality factor from 0.1 to 0.9, the byte hit rate improves 9% ~ 22% and essentially reaches the performance of network-wide flooding. The average hops metric has 7% ~ 19% improvement. In terms of byte hit rate, the difference of all three strategies becomes smaller as the locality factor increases but dynamic flooding consistently outperforms static one by at least 7.5%. Meanwhile, the cost of all three strategies almost remain unchanged. The reason is that content availability and local topological property are the determinant factors of the cost (due to being a function of p and r) in static and dynamic flooding respectively, neither will be affected by changing the locality factor.

6.5 Summary and Discussion

Our results indicate that dynamic flooding is more effective on the networks of heterogeneous topological structure, and most of the gains come from sparse areas wherein local growth rate is low, namely at network edges. The optimal flooding radius in a dense area is small and nodes suffer from high flooding cost. On realistic networks for which the evaluations further take the spatial locality of content into account, dynamic flooding is consistently superior to static one even without prior knowledge on content availability, and quickly approaches the byte hit rate of network-wide flooding but with much smaller cost in control messages.

7. RELATED WORK

We can categorize the literature into two as *resolution-based discovery* and *routing-based discovery*.

Resolution-based discovery [2–8] provide deterministic solutions, i.e., at least one copy will be found as long as the content is stored within the network. Therefore, such solutions either require complete knowledge on content distribution and network topology [2, 4, 5] or utilize hash-based content addressing [3, 6–8]. Essentially, demands and supplies meet at *rendezvous points* (the actual name differs depending on specific architecture). The rendezvous point either returns a locator or copy of the requested content [2, 3], or redirects the request to one content provider [5, 7], or constructs a distribution topology [4] depending on an actual design. Resolution-based solutions can reach high success rate but have to maintain the states of content distribution in a network hence are confronted with scalability issue when dealing with large and highly dynamic content demands.

Routing-based discovery [1, 9, 10, 32] usually only provides opportunistic solutions, i.e., content will be found with certain probability. The chances of discovery can be improved by either collaborating with nearby nodes [10, 35] or exploring a network via flooding [9, 15, 16, 32, 36]. Both introduce extra traffic overhead. [35] propose using Bloomfilters to exchange information on content availability to improve caching performance. [15, 16, 37] empirically showed that opportunistic flooding can improve content discovery and delivery, also reduce the states maintained in a network. [9, 36, 38, 41] showed that flooding is especially preferred in an unreliable environment to compensate for potential message loss. Empirically or analytically, all [9, 16, 32, 38]

attested that naive flooding is hardly viable in practice, the scope needs to be regulated carefully to reduce the cost.

Scoped-flooding also relates to *Gossip Protocol* and *Expanding Ring Search (ERS)*. Gossip protocol has been shown as a simple, robust, and scalable solution on large distributed systems [11]. ERS is supported in reactive routing protocols in MANET such as DSR [39] and AODV [40]. Since nodes cache routes information like content caching in ICN, ERS is rather similar to scoped-flooding. Prior work [19–21] showed that the radius is very small in practice. Regarding the neighbourhood growth model, besides [22, 24, 25], another important line of research is *expander graph* [26]. In general, the advances in graph theory has improved our understanding on network graphs and laid the foundation of this work. Nonetheless, as far as we can tell, none carefully analysed the scoped-flooding for content discovery from network topology perspective, not to mention examining the distribution of gains and improvements within a network.

8. CONCLUSION

This paper aims to comprehend scoped-flooding for content discovery in ICN. Using the proposed ring model, we studied the functional relation between the neighbourhood growth and flooding radius, based on which we derived the optimal search radius. Both our theoretical analysis and empirical evaluations suggest that due to the exponential growth of neighbourhood size, the optimal flooding radius is usually very small (i.e., a couple of hops). Most of the gains of flooding come from the sparse area at the network edge where the neighbourhood growth rate is low. To certain extent, our results justify the rationale of deploying caches at network edge from content discovery perspective. Dynamic flooding is consistently superior to static one, especially on scale-free networks. With strong spatial locality, the performance of dynamic flooding quickly converges to network-wide flooding but with much smaller cost. However, we acknowledge that the following aspects need further investigation in future: (1) current neighbourhood growth model does not take clustering coefficient into account which leads to overestimation in small networks. (2) Our utility model assumes sub-linear gain and linear cost which requires further reality checks. (3) Other in-network caching algorithms can be more effective than simple LRU and a thorough comparison is definitely needed to gain a deeper understanding.

9. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union’s (EU) Horizon 2020 research and innovation programme under grant agreement No.645124 (Action full title: Universal, mobile-centric and opportunistic communications architecture, Action Acronym: UMOBILE), and No.644663 (Action full title: architectuRe for an Internet For Everybody, Action Acronym: RIFE) This paper reflects only the authors’ views and the Community is not liable for any use that may be made of the information contained therein.

10. REFERENCES

- [1] V. Jacobson, et al., “Networking named content,” in *Proceedings of the 5th ACM Conext*. ACM, 2009.
- [2] C. Dannewitz, et al., “Netinf - an information-centric networking architecture,” *Comput. Commun.*, 2013.

- [3] M. D'Ambrosio, et al., "MDHT: a hierarchical name resolution service for information-centric networks," in *ACM SIGCOMM workshop on ICN*, ACM, 2011.
- [4] D. Trossen, et al., "PURSUIT conceptual architecture: principles, patterns and sub-components descriptions," 2011.
- [5] T. Koponen, et al., "A data-oriented (and beyond) network architecture," *Comput. Commun. Rev.*, 2007.
- [6] D. G. Thaler, et al., "Using name-based mappings to increase hit rates," *Transactions on Networking*, 1998.
- [7] L. Wang, et al., "MobiCCN: Mobility support with greedy routing in Content-Centric Networks," in *Globecom*, 2013.
- [8] S. Roos, et al., "Enhancing Compact Routing in CCN with Prefix Embedding and Topology-aware Hashing," in *Mobicom workshop on MobiArch'14*, 2014.
- [9] M. Varvello, et al., "On the design of content-centric MANETs," in *WONS*, 2011.
- [10] W. Wong, et al., "Neighborhood search and admission control in cooperative caching networks," in *Globecom*, *IEEE*, 2012.
- [11] P. Eugster, et al., "Epidemic information dissemination in distributed systems," *Computer*, 2004.
- [12] S. Traverso, et al., "Temporal locality in today's content caching: why it matters and how to model it," *Computer Communication Review*, ACM, 2013.
- [13] E. Baccelli, et al., "Information centric networking in the IoT: Experiments with NDN in the wild," in *ICN'14*.
- [14] PARC, "CCNx 1.0 protocol specifications," 2014.
- [15] M. Badov, et al., "Congestion-aware caching and search in information-centric networks," in *ICN'14*.
- [16] R. Chiocchetti, et al., "Exploit the known or explore the unknown? Hamlet-like doubts in ICN," in *ICN'12*.
- [17] J. Garcia-Luna-Aceves, "Name-based content routing in information centric networks using distance information," in *ICN'14*, ACM, 2014.
- [18] Q. Lv, et al., "Search and replication in unstructured peer-to-peer networks," in *ICS'02*, ACM, 2002.
- [19] H. Barjini, et al., "Shortcoming, problems and analytical comparison for flooding-based search techniques in unstructured P2P networks," *P2P Netw. and Appl.*, 2012.
- [20] J. Hassan, et al., "Optimising expanding ring search for multi-hop wireless networks," in *GLOBECOM'04*.
- [21] J. Deng, et al., "On search sets of expanding ring search in wireless networks," *Ad Hoc Netw.*, 2008.
- [22] M. Newman, "Random graphs as models of networks", Wiley, 2003.
- [23] S. Bornholdt, et al., "Handbook of graphs and networks", Wiley, 2003.
- [24] M. Molloy, et al., "A critical point for random graphs with a given degree sequence," *Random Struct. Algorithms*, 1995.
- [25] S. L. Feld, "Why your friends have more friends than you do," *The American Journal of Sociology*, 1991.
- [26] S. Hoory, et al., "Expander graphs and their applications", Bulletin of the American Mathematical Society, 2006.
- [27] N. Spring, et al., "Measuring ISP topologies with Rocketfuel," in *SIGCOMM*, ACM, 2002.
- [28] D. Vega, et al., "Topology patterns of a community network: Guifi.net," in *WiMob*, *IEEE*, 2012.
- [29] R. Hekmat, et al., "Connectivity in wireless ad-hoc networks with a log-normal radio model," *Mob. Netw. Appl.*, 2006.
- [30] E. Yoneki, et al., "Distinct types of hubs in human dynamic networks," in *Social Network Systems*, 2008.
- [31] A. Clauset, et al., "Power-law distributions in empirical data," *SIAM review*, 2009.
- [32] E. Hytiä, et al., "Searching a needle in (linear) opportunistic networks," in *MSWiM '14*, ACM, 2014.
- [33] M. Cha, et al., "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *IMC'07*, ACM, 2007.
- [34] A. Dabirmoghaddam, et al., "Understanding optimal caching and opportunistic caching at "the edge" of information-centric networks," in *ICN'14*, 2014.
- [35] M. Lee, et al., "Content discovery for information -centric networking," *Computer Networks*, 2014.
- [36] N. Chang, et al., "Controlled flooding search in a large network," *Transactions on Networking*, 2007.
- [37] G. Rossini, et al., "Coupling caching and forwarding: Benefits, analysis, and implementation," in *ICN'14*.
- [38] C. Liu, et al., "An optimal probabilistic forwarding protocol in delay tolerant networks," in *MobiHoc'09*.
- [39] J. Broch, et al., "The dynamic source routing protocol for mobile ad hoc networks," *IETF draft*, 1999.
- [40] C. Perkins, et al., "Ad-hoc on-demand distance vector routing," in *WMCSA '99*, 1999.
- [41] A. Sathiaselan, et al., "SCANDEX: Service Centric Networking for Challenged Decentralised Networks," in *Mobisys workshop on DIYNet*, 2015.
- [42] L. Wang, et al., "Effects of Cooperation Policy and Network Topology on Performance of In-Network Caching," in *Communications Letters, IEEE*, 2014.