

Revista Eletrônica de Sistemas de Informação

ISSN 1677-3071

v. 16, n. 2

mai-ago 2017 - Edição temática sobre Computação Urbana

DOI: <https://doi.org/10.21529/RESI.2017.1602>

Sumário

Computação Urbana

O PAPEL DA UNIVERSIDADE NA CONSTRUÇÃO DE CIDADES INTELIGENTES E HUMANAS

Ana Regia de Mendonca Neves, Kaê U. Sarmanho, Bianchi S. Meiguins

[doi> 10.21529/RESI.2017.1602001](https://doi.org/10.21529/RESI.2017.1602001)

PROPOSTA DE UM FRAMEWORK BASEADO EM MINERAÇÃO DE DADOS PARA REDES 5G

Carlos Renato Storck, Edwaldo Araújo Sales, Luis Enrique Zárate, Fátima de L. P. D. Figueiredo

[doi> 10.21529/RESI.2017.1602002](https://doi.org/10.21529/RESI.2017.1602002)

SERVIÇOS DE EMERGÊNCIA EM CIDADES INTELIGENTES: O PROBLEMA DE ACIONAMENTO DE UNIDADES MÓVEIS

Sediane Carmem Lunardi Hernandes, Alcides Calsavara, Marcelo Eduardo Pellenz, Luiz Augusto de Paula Lima Júnior

[doi> 10.21529/RESI.2017.1602003](https://doi.org/10.21529/RESI.2017.1602003)

USANDO O CLASSIFICADOR NAIVE BAYES PARA GERAÇÃO DE ALERTAS DE RISCO DE ÓBITO INFANTIL

Cristiano Lima da Silva, Joyce Quintino Alves, Oton Crispim Braga, José Wellington Pereira Júnior, Luiz Odorico Monteiro de Andrade, Antônio Mauro Barbosa de Oliveira

[doi> 10.21529/RESI.2017.1602004](https://doi.org/10.21529/RESI.2017.1602004)



Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

This journal is (and has always been) electronic in order to be more environmentally friendly. Now, it is desktop edited in a single column to be easier to read on the screen. However, if you wish to print this paper, be aware that it uses Eco Sans, a printing font that reduces the amount of required ink.

PROPOSTA DE UM *FRAMEWORK* BASEADO EM MINERAÇÃO DE DADOS PARA REDES 5G

DATA MINING BASED FRAMEWORK PROPOSAL FOR 5G NETWORKS

(artigo submetido em agosto de 2017)

Carlos Renato Storck

Doutorando em Informática pela Pontifícia Univ. Católica de Minas Gerais (PUC-MG)
Professor do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
storck@cefetmg.br

Luis Enrique Zárate

Doutor em Eng. Metalúrgica e de Minas pela Univ. Federal de Minas Gerais (UFMG)
Professor do Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica de Minas Gerais (PUC-MG)
zarate@pucminas.br

Edwaldo Araújo Sales

Mestrando em Informática pela Pontifícia Univ. Católica de Minas Gerais (PUC-MG)
ed.araujo@outlook.com

Fátima de L. P. D. Figueiredo

Doutora em Ciência da Computação pela Univ. Federal de Minas Gerais (UFMG)
Professora do Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica de Minas Gerais (PUC-MG)
fatimafig@pucminas.br

RESUMO

Cidades inteligentes vêm ganhando, cada vez mais, notoriedade. Através delas, a população pode ter melhores serviços e qualidade de vida urbana. Com as futuras redes de celulares de quinta geração (5G) será possível coletar dados por meio de diversas fontes espalhadas pela cidade, tais como sensores, dispositivos móveis, redes veiculares e de telefonia, dentre outras. Nesse cenário, haverá a necessidade de análise de grandes volumes de dados, com o objetivo de extrair conhecimento e informação útil para o planejamento inteligente e dinâmico. Este artigo apresenta uma proposta de *framework* baseado em mineração de dados para redes 5G, denominado *Urban Computing Framework in 5G Networks* (CoUrbF5G). Padrões reais de uma rede de telefonia móvel são encontrados e analisados, aplicando técnicas de mineração de dados, em conjunto com métodos auxiliares na condução de processos como *Knowledge Discovery in Databases* (KDD) e *Big Data*.

Palavras-chave: Redes 5G; mineração de dados; computação urbana.

ABSTRACT

Smart cities are gaining, increasingly, notoriety. Through them, the population can have better services and urban quality of life. With the future fifth generation cellular (5G) networks will be possible to collect data from various sources throughout the city, such as sensors and actuators, mobile devices, mobile and telephone networks among others. In this scenario, it will be necessary to analyze large volumes of data, in order to get knowledge and useful information for a dynamic and clever planning. This article presents a data mining based framework proposal for 5G networks, named Urban Computing Framework in 5G Networks (CoUrbF5G). Real standards of a mobile network are found and analyzed in the proposal by applying data mining techniques along with ancillary methods on the conducting of Data Science processes, such as Knowledge Discovery in Databases (KDD) and Big Data.

Key-words: 5G networks; data mining; urban computing.

1 INTRODUÇÃO

O planejamento urbano em cidades inteligentes requer a utilização de técnicas e mecanismos de Computação Urbana (CoUrb) em ambientes de *Big Data*. Tais mecanismos ou ambientes permitem a entrega de melhores serviços para a cidade, melhorando, conseqüentemente, a qualidade de vida dos cidadãos (ZHENG *et al.*, 2014).

Dessa maneira, para entender os fenômenos urbanos, realizar previsões e superar os problemas que as cidades enfrentam, tais como congestionamentos, consumo de energia, poluição do ar, entre outros, são necessários processamento e análise de diversos dados, através da CoUrb. Quanto à coleta de dados, as cidades inteligentes chegarão a conectar bilhões de dispositivos e sensores. Em tal cenário, os dados são coletados por diversas fontes espalhadas pela cidade, tais como sensores de temperatura, umidade, energia, presença, redes veiculares e de telefonia, dentre outros.

Para o processamento e análise de diversos dados, justifica-se o uso de técnicas como a mineração de dados como apoio para a descoberta de padrões preditivos não óbvios, por exemplo, padrões de mobilidade urbana, comportamento dos usuários, uso da rede, dentre outros. A mineração de padrões se justifica pela importância de encontrar padrões interessantes, inesperados e úteis de um domínio, objetivando construir uma base de conhecimento para auxílio na tomada de decisão. No contexto da Ciência de Dados (*Data Science*), o processo KDD é realizado em diversas etapas: seleção de dados, pré-processamento, transformação, mineração de dados, validação e interpretação (FAYYAD *et al.*, 1996).

Neste trabalho, é aplicado um processo de KDD em ambiente de grande volume de dados. Caracterizou-se o comportamento de uma rede de telefonia móvel utilizando uma base com dados reais, através de técnicas de mineração de dados em conjunto com um método auxiliar na condução desse processo, para subsidiar o *framework* proposto pela abordagem do *Urban Computing Driven Data Mining* (CoUrbD2M).

O CoUrbD2M possibilita a mineração de dados orientada à computação urbana, permitindo a descoberta de padrões de condições da rede e fornecendo entradas para o *framework* CoUrbF5G. Este, por sua vez, permite que a rede se oriente pelo contexto, aprenda com a experiência e execute ações necessárias para alocação apropriada de recursos em conjunto com a tomada de decisões dinâmicas de planejamento urbano, análise de regiões funcionais e detecção de anomalias.

Este trabalho se desenvolve da seguinte forma: a Seção 2 apresenta os principais conceitos técnicos e aponta os trabalhos de *Data Mining* relacionados ao estudo de *Big Data* e das redes 5G; a Seção 3 descreve a metodologia; a Seção 4 apresenta e discute os resultados da pesquisa; e a Seção 5 discorre as conclusões e proposições a trabalhos futuros.

2 PRINCIPAIS CONCEITOS E TRABALHOS RELACIONADOS

Nesta seção são apresentados os principais conceitos e trabalhos relacionados para a compreensão de como a mineração de dados pode ser aplicada em cenários de *Big Data* e redes 5G com o objetivo de permitir a tomada de melhores decisões de alocação de recursos e planejamento.

2.1 DATA MINING E BIG DATA

A mineração de dados espaço-temporais é o processo de descoberta de padrões em dados espaciais associados com o tempo. Essa é uma área de pesquisa aplicada em redes móveis e pode ser associada com a mineração de realidade, que trata o estudo do comportamento social humano com base em redes móveis (BAYIR *et al.*, 2009; EAGLE *et al.*, 2009; AYDIN; ANGRYK, 2016).

Há tarefas mais comuns utilizadas na mineração de dados espaço-temporais. A primeira tarefa relatada é o agrupamento espacial, considerando a análise da área geográfica para o qual os autores adotaram o algoritmo *k-means*. A segunda é a análise de *outliers*, que consiste em observar valores distantes ou fora do padrão de uma amostra de uma população. A terceira tarefa apontada é o uso de modelos preditivos para identificar padrões e prever valores de algum evento com base em dados históricos. A última tarefa relatada é a busca por correlações e objetos (registros) que possuem uma alta probabilidade de ocorrência, por exemplo, encontrar subconjuntos agrupados de tipos de eventos (MORE; LINGAM, 2015).

Existe uma proposta de arquitetura de mineração em ambientes de *Big Data* para dados massivos, produzidos pela Internet das Coisas (IoT). Aplicações de *e-commerce*, indústria, saúde e governo exigirão algoritmos adaptados de classificação, agrupamento (*clustering*), análise de associação, análise de séries temporais e *outliers* (CHEN *et al.*, 2015a). A coleta de dados sensoriais em larga escala, a partir de dispositivos móveis, para a compreensão de congestionamentos, níveis de poluição e ruído, entre outros, juntamente com o estado da arte e as direções correntes para a mineração de fluxo de dados em ambientes ubíquos são abordados por um conjunto de ferramentas denominado *Open Mobile Miner* (GABER *et al.*, 2014).

Também foi encontrada uma proposta para melhorar o desempenho e a eficiência da rede. Os dados são coletados através da sinalização, tráfego e ambiente. O fluxograma da otimização da rede inteligente baseada em dados de sinalização, apresentado na Figura 1, é composto pelas seguintes etapas: limpeza de dados, seleção de dados e integração, armazenamento, mineração de dados e soluções de otimização da rede. Entre as características do modelo do fluxograma, destaca-se a descoberta de problemas da rede através dos indicadores de desempenho. É gerada uma base de soluções de otimização da rede com as melhores práticas com base na experiência de otimização para os problemas identificados (Liu *et al.*, 2015).

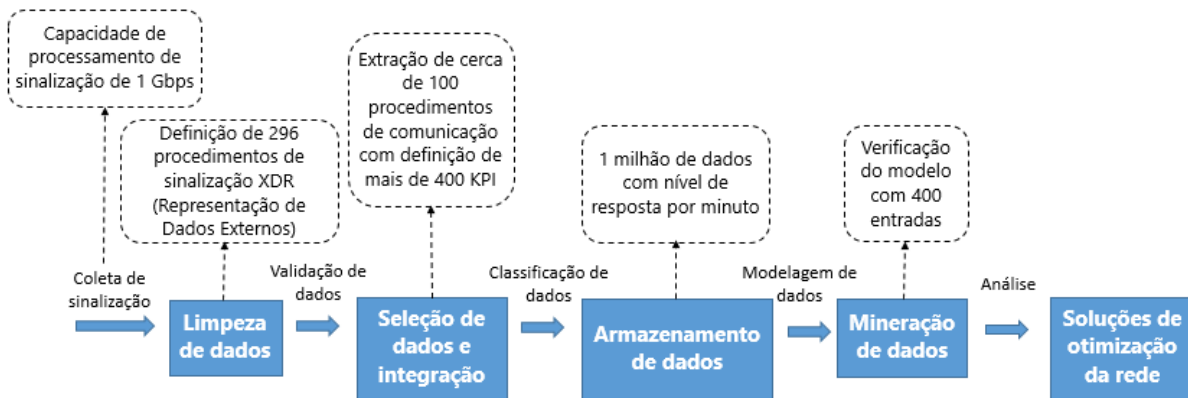


Figura 1. Fluxograma da otimização da rede inteligente baseada em dados de sinalização

Fonte: adaptada de Liu *et al.* (2015)

No trabalho de Liu *et al.* (2015), os autores também apresentam o paradigma de operação da rede baseada em *Big Data*, apresentado pela Figura 2, que inclui a coleta de informações de usuários, permitindo gerenciar e configurar funções de acordo com o contexto. Nos estudos He *et al.* (2016) também são apresentadas arquiteturas para aplicação de *Big Data* em redes celulares móveis, incluindo exemplos de dados de sinalização, tráfego, localização, ondas de rádio e dados heterogêneos.

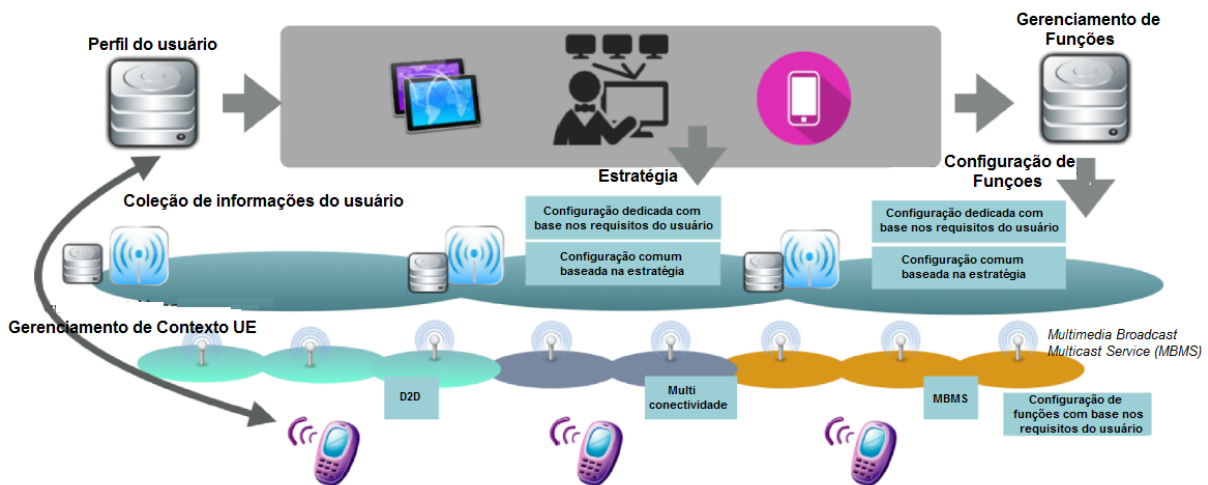


Figura 2. Paradigma de operação da rede baseada em *Big Data*

Fonte: adaptada de Liu *et al.* (2015)

Uma arquitetura da rede sem fio, em nuvem, com centros de *Big Data* é proposta por Chen *et al.* (2015b) é apresentada na Figura 3. A ideia central é alocação dinâmica de recursos baseada em *Big Data*, analisando a rede sob diferentes cenários de tráfego. Para QoS (*Quality of Service*), a otimização com base em dados de tráfego de rede sem fio fica responsável pela cobertura, realocação de recursos e detecção de falhas. Para QoE (*Quality of Experience*), a otimização com base em dados no comportamento fica responsável pela otimização de dispositivos e sua localização.

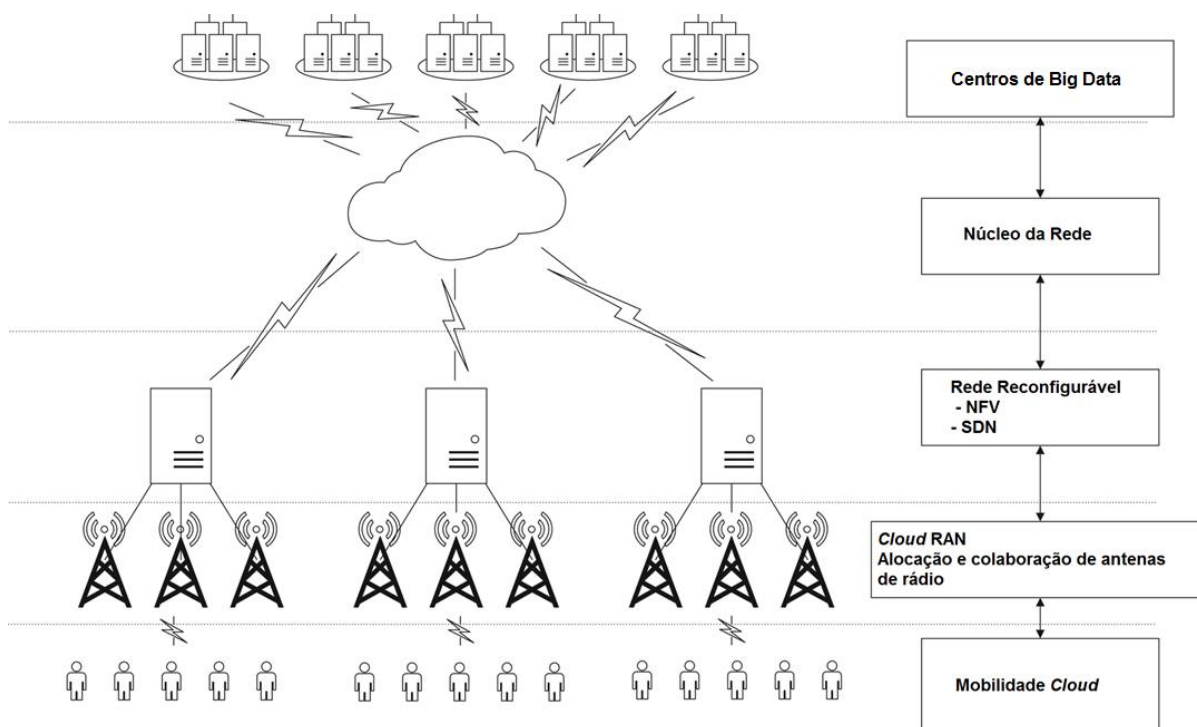


Figura 3. Arquitetura da rede sem fio na nuvem com centros de *Big Data*

Fonte: adaptada de Chen *et al.* (2015b)

Em se tratando de *Big Data*, como incentivo à comunidade acadêmica para superação dos desafios, Barlacchi *et al.* (2015) apresentam um conjunto de bases de dados da vida urbana italiana. Uma dessas bases de dados (georreferenciada e anonimizada), que serviu de base para este trabalho, é disponibilizada no formato ODbL (*Open Database License*), pela companhia Telecom Itália, maior empresa de telecomunicações desse país, juntamente com outras instituições, através do “*Telecom Italia Big Data Challenge*”¹. O *dataset* apresenta chamadas telefônicas, serviços SMS (*Short Message System*) e uso de Internet por horários extraídos de registros de detalhamento da chamada (CDR). Esses registros são extremamente importantes para operadoras, pois permitem administrar e faturar os serviços utilizados na rede. Na literatura são encontrados vários trabalhos que usam CDR para analisar o comportamento da rede.

2.2 REDES 5G

Redes 5G podem ser definidas como redes ubíquas de alta largura de banda, que causarão uma verdadeira revolução nas redes móveis. Elas se encontram, atualmente, em fase de pesquisas e testes, possuindo ainda vários desafios. Entre as características e requisitos exigidos por essas redes estão o aumento no *throughput*, latência inferior a 1ms, ultra confiabilidade, cobertura adequada para uma experiência perfeita, bateria dos dispositivos com vida útil 10 vezes maior e uso de todos os espectros. A plataforma necessária para as redes 5G das operadoras, a partir de

¹ <http://www.telecomitalia.com/tit/en/bigdatachallenge/contest.html>

2020, vai precisar de uma arquitetura definida por *software*, capaz de ser executada em um sistema integrado sem fio e óptico, permitindo unificar conexões, ampliar segurança, mobilidade e gerenciamento do encaminhamento.

Uma visão da arquitetura 5G é proposta por Agyapong *et al.* (2014). Essa arquitetura é formada por duas camadas: uma Rede de Acesso de Rádio (RAN) e uma rede em nuvem. Nessa arquitetura serão integradas as pico células, múltiplos transmissores e receptores, divisão do plano de usuário e controle, SDN (*Software-Defined Networking*) e NFV (*Architecture and Network Function Virtualization*).

Ainda de acordo com Agyapong *et al.* (2014), há desafios e princípios de projeto para redes 5G. Entre as propostas citadas, a inteligência de rede orientada em *Big Data* é apontada para superar desafios como custo e QoE. O objetivo seria adotar um agente inteligente de rede para gerenciar a qualidade de serviço, o roteamento, a mobilidade e a alocação de recursos de forma dinâmica. Para tanto, a descoberta de conhecimento e mineração de dados em ambientes de *Big Data* é essencial.

Para uma arquitetura 5G ágil e flexível será necessário separar o plano de dados do plano de controle e implementar funções de rede virtualizadas com os fluxos de dados orquestrados por meio de programação. As redes 5G, provavelmente, terão maior volume de tráfego e número de dispositivos com requisitos diversificados e melhor qualidade de experiência do usuário. A melhoria da eficiência operacional da rede se dará através do uso de SDN e NFV, mas ainda é esperado o uso de tecnologias, como C-RAN (*Cloud-RAN*) e SON (*Self-Organizing Network*) (ITU, 2015).

Em Malik *et al.* (2015), são discutidas técnicas de QoS para redes sem fio atuais e emergentes. Para a ciência do contexto é relatada uma forma de melhorar tanto a qualidade de serviço, como a qualidade de experiência. A computação ciente de contexto permite verificar alterações no ambiente e adaptar o sistema ao novo estado, enriquecendo a comunicação entre pessoas e dispositivos computacionais. Como o comportamento do tráfego varia no tempo, a análise do comportamento é sugerida como solução para prever futuros padrões de tráfego; para que se possa, assim, prever a mobilidade e melhorar a alocação de recursos e transmissão de dados, bem como permitir adoção de estratégias específicas para a rede em determinado momento. O ambiente CWN (*Cognitive Wireless Network*) é proposto, por esses autores, com um motor cognitivo embutido, que observa as condições do ambiente, realiza sua análise, consulta políticas e objetivos da rede, além de realizar seu planejamento, aprender com a experiência e atuar da melhor forma possível. Mas, além da reconfiguração em nível de rede, através de SDN, é ainda sugerida a implementação de planos de conhecimento.

3 COURBF5G E COURBD2M

Neste trabalho, CoUrbF5G é a proposta de *framework* que se baseia em rede cognitiva para redes 5G, apresentado pela Figura 4. Com os dados coletados pela rede 5G, através de diversas fontes espalhadas pela cidade e armazenados em *data centers*, o *framework* possui como entrada a abordagem de mineração de dados orientada à computação urbana, que será descrita a seguir, o tornando capaz de realizar um adequado planejamento urbano e verificar padrões ou previsões de demandas futuras das cidades, por exemplo, uma grande quantidade de pessoas presentes em um determinado evento, como *shows* ou jogos em estádios.

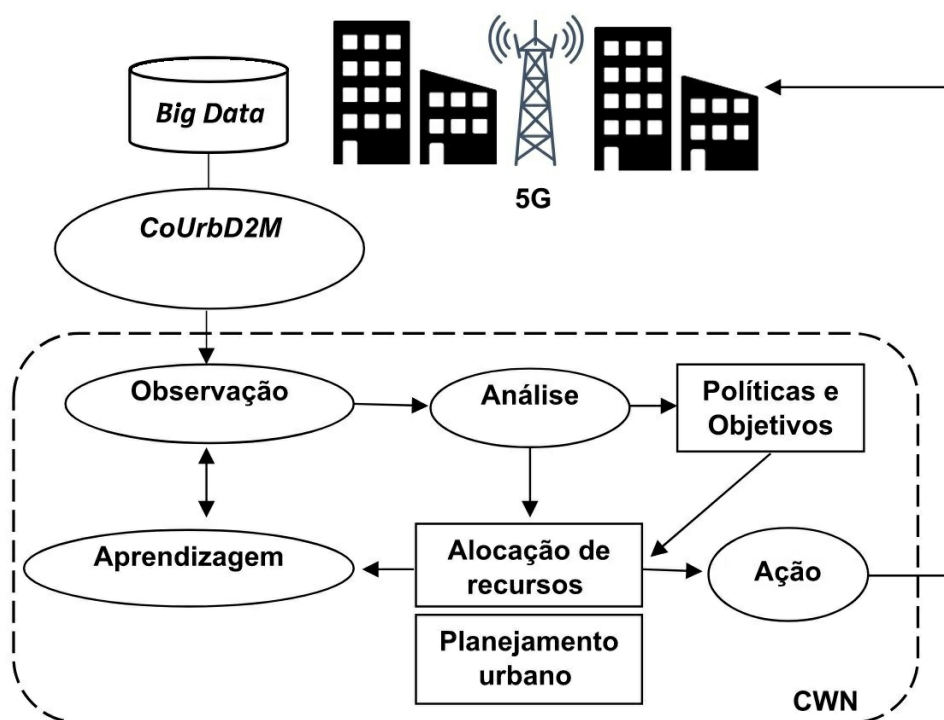


Figura 4. Representação do *framework* CoUrbF5G

Fonte: elaborada pelos autores

Conforme apresentado na Figura 4, o *framework* CoUrbF5G é formado por componentes diversos. O CoUrbF5G inicia-se com a coleta de diversos tipos de dados da computação urbana, através da infraestrutura das futuras redes 5G, que serão armazenadas em grandes bases de dados hospedadas em *data centers*. Em seguida, a mineração de dados orientada à computação urbana é executada, entregando os resultados para o ambiente CWN. Nesse ambiente, o módulo de observação verifica o comportamento ciente de contexto, com base nos resultados da CoUrbD2M, e o módulo de análise verifica as condições atuais da rede. Observado o comportamento e realizada a análise, são consultados os parâmetros desejados, através do módulo de políticas e objetivos. Os módulos de alocação de recursos e planejamento urbano, então, serão acionados, realizando as devidas mudanças dinâmicas do sistema e da rede pelo módulo de ação, capacitando o sistema para aprender com a experiência,

através do módulo de aprendizagem, que poderá alterar os parâmetros do módulo de observação.

Como módulo de entrada para subsidiar o *framework*, o CoUrbD2M é uma proposta para mineração de dados. Como a quantidade de dados gerados será muito grande, a proposta visa melhor interpretação e análise dos dados. CoUrbD2M foi desenvolvida tendo como base as diversas etapas propostas pelo método PICTOREA, descrito por Montevecchi (2012). O método serviu de base para auxílio na condução e integração da percepção humana no processo de KDD. A Figura 5 mostra o *workflow* do método, composto por treze etapas no seu fluxo principal e caracterizado pelo próprio fluxo mantendo a concisão das etapas do projeto. As etapas envolvem desde o entendimento do problema até a validação e testes dos padrões encontrados para análise dos especialistas. Para o contexto deste trabalho, CoUrbD2M foi limitada à aplicação de métodos padrão de mineração de dados.

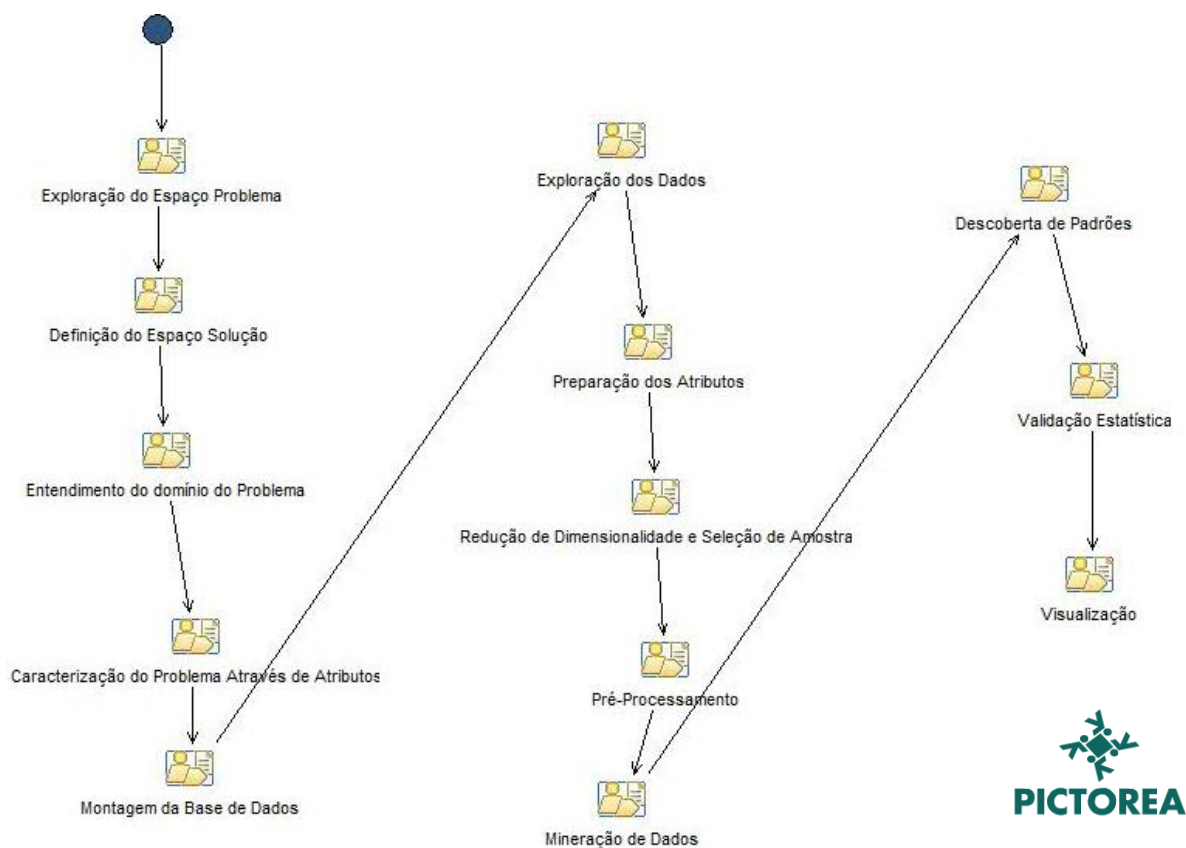


Figura 5. *Workflow* do método PICTOREA

Fonte: PICTOREA (2016)

Sendo assim, adotou-se um conjunto de dados reais de telefonia móvel, coletado durante uma semana em Milão e na região de Trentino, durante o mês de novembro de 2013. Os dados foram disponibilizados pela companhia Telecom Itália, no formato ODbL, para a competição *Big Data Challenge*, conforme mencionado na seção anterior. A base de dados

é composta por mais de 15 milhões de registros, tendo cada registro a atribuição da antena, ou Estação Rádio Base (ERB), e a contabilização por tipo de serviço e por horário. A base de dados é formada pelos atributos de recebimento e realização de chamadas telefônicas, recebimento e envio de SMS e uso de Internet, sendo válido para realizar uma análise preliminar de planejamento das futuras redes 5G. A ferramenta Weka, composta por uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados, de código aberto e desenvolvida pela Universidade de Waikato na Nova Zelândia, foi usada para analisar o conjunto de dados (WEKA, 2016).

A primeira etapa do CoUrbD2M é a exploração do espaço do problema para conhecimento de cada domínio aplicado na computação urbana. Observa-se que diversos domínios podem ser criados a partir dos problemas enfrentados pelas cidades, por exemplo, a verificação de mobilidade para superar os desafios de congestionamentos urbanos. Para alcançar o objetivo desta etapa, foi utilizada uma matriz de problemas *pairwise*, apresentada na Tabela 1, atribuindo peso 0,5 para importância do problema e peso 0,25 para facilidade e retorno. Os problemas-alvo levantados foram: a investigação dos horários de pico e demanda por antena, bem como a similaridade entre os serviços prestados, objetivando um correto planejamento futuro.

Tabela 1. Matriz de problemas *Pairwise*

Problema	Importância	Facilidade	Retorno	Total
Horários de pico	3	2	3	2,75
Demanda por antena	2	3	2	2,25
Similaridade entre serviços	2	3	1	2,00

Fonte: elaborada pelos autores

A próxima etapa sugerida é a definição do espaço solução. Nessa etapa, foram definidas as técnicas de mineração de dados e visualização, considerando os resultados esperados. Adotou-se como técnica de mineração para atendimento das expectativas, a técnica de agrupamento.

Para compreensão do domínio do problema e sua caracterização, foram verificadas quais características enfatizariam o conhecimento útil e não óbvio, podendo, então, ser avaliado usando os conceitos e aspectos *Domain Driven Data Mining* (D3M), onde cada etapa do processo KDD deve ser acompanhada e validada por um especialista de domínio. A caracterização do problema através de atributos identificou que os atributos *datetime*, *cell-ID*, *sms*, *call* e *internet* são os mais relevantes, considerando o problema definido e que será analisado.

Uma das etapas que mais consome tempo e recursos é a montagem da base de dados. Neste trabalho, a montagem da base de dados se deu no formato de arquivo *Comma-separated values* (CSV) através de *scripts* na linguagem C#, verificando sua dimensionalidade. Foram verificadas a

consistência e coerência dos atributos das instâncias, a presença de poluição nos dados e a integridade. A base final foi considerada representativa para descoberta de conhecimento.

Após a montagem da base de dados, a análise da representatividade estatística foi realizada através da etapa da exploração de dados. Também procurou-se obter a análise de correlação para avaliar o grau de relacionamento entre as variáveis.

Na etapa de preparação dos atributos, foram verificados os valores ausentes e a análise de *outliers* presentes na base de dados. Para valores ausentes, optou-se por eliminar todos os registros encontrados.

A redução da dimensionalidade e seleção de amostras é de suma importância para uma boa representação da descoberta de conhecimento, sendo necessário avaliar os atributos para esta execução. Os atributos *sms-in* e *sms-out* foram unidos em único atributo *sms*, por se considerar a contagem independente de envio ou recebimento. O mesmo ocorreu com os atributos *call-in* e *call-out*. Neste trabalho, foram utilizadas duas amostras representativas de instâncias do banco de dados, o que resultou em um primeiro subconjunto de dados formado por 499.999 instâncias, e um segundo subconjunto contendo 2.008.105 instâncias, sendo que a base original continha 15.085.579 registros.

Os resultados estão apresentados na próxima seção, permitindo a avaliação do especialista de domínio da computação urbana, conforme propõe o método de base PICTOREA.

4 RESULTADOS

Com o intuito de verificar a viabilidade do módulo de mineração orientada à computação urbana, os resultados da análise obtida nos experimentos são apresentados nesta Seção, os quais serão tratados futuramente pela implementação do CoUrbF5G. Conforme descrito na Seção 3, o *dataset* foi explorado e realizou-se o pré-processamento através da ferramenta Weka.

4.1 CLUSTERING

Para a etapa de agrupamento, a análise de *outliers*, que representam valores fora do padrão dos dados é essencial, uma vez que os *outliers* podem ser causadores de distorções nas formações de grupos. Com esse intuito, para análise de *outliers*, aplicou-se a média com dois desvios padrão, sendo encontrados 60.313 registros considerados fora do padrão, o que representa 12,06% do primeiro subconjunto de dados. Em um segundo experimento, com um subconjunto de dados de 2.008.105 registros, foram encontrados 97.840 registros considerados *outliers*, o que representa 4,87% deste segundo subconjunto de dados.

Para análise de agrupamento de dados, foi usada primeiramente a técnica de particionamento *k-means*, que é considerada uma técnica de aplicação simples, relativamente escalável e eficiente para grandes ban-

cos de dados, com o objetivo de identificar os registros similares. Uma proposta de análise de *clusters* é encontrada em (SERRA; ZÁRATE, 2015). Foram adotados dois (Tabela 2) e cinco *clusters* (Tabela 3), $k=2$ e $k=5$, respectivamente. Sendo a quantidade de *clusters* desejados, k , fornecidos previamente pelo usuário, os quais serão representados por seus centroides. A adoção de cinco *clusters* como medida de avaliação do agrupamento apresentou uma boa representação em relação ao número de *clusters* pelo tamanho da amostra, sendo esse valor também empregado no trabalho de More e Lingam (2015). Na Tabela 3, verifica-se que dois grupos (*clusters* 2 e 3) apresentaram um maior tamanho, o que significa que os serviços prestados em determinadas antenas e por horário seguem um padrão similar.

Tabela 2. *k-means* com dois *clusters*

<i>Cluster</i>	Instâncias	Resultado
0	267.686	54%
1	232.313	46%

Fonte: elaborada pelos autores

Tabela 3. *k-means* com cinco *clusters*

<i>Cluster</i>	Instâncias	Resultado
0	75.021	15%
1	65.283	13%
2	118.672	24%
3	187.536	37%
4	53.487	11%

Fonte: elaborada pelos autores

Em seguida, o algoritmo EM (*Expectation Maximization*), que é um método baseado em modelos probabilísticos, foi executado para fins de comparação com *k-means*, conforme apresentado na Tabela 4. O objetivo da comparação era verificar se havia diferença significativa na avaliação dos agrupamentos, empregando outro algoritmo. Contudo, assim como o *k-means*, verifica-se que dois grupos (*clusters* 0 e 3) apresentaram um maior tamanho, o que significa que os serviços prestados em determinadas antenas e por horário seguem um padrão similar.

Tabela 4. *Expectation Maximization* com cinco *clusters*

<i>Cluster</i>	Instâncias	Resultado
0	145.396	29%
1	92.946	19%
2	80.155	16%
3	114.593	23%
4	66.909	13%

Fonte: elaborada pelos autores

Com o objetivo de verificar as demandas por antena, a Figura 6 mostra a contabilização de chamadas telefônicas por antena, permitindo identificar regiões com uso intenso na rede. As cores representam as chamadas telefônicas agrupadas em cinco *clusters*. É possível perceber, na figura, que as antenas entre 500 e 1000 possuem maior demanda em determinados períodos de tempo. Relacionando-se a demanda de chamadas com o mapa real, é constatado que em Milão, a área de cobertura da Universidade de Bocconi possui menor demanda do que a Catedral de Milão, que é o centro da cidade e considerada a atração turística mais importante. O distrito de Navigli, região com vida noturna agitada, também apresenta uma demanda alta. A região com menor demanda é a área de Bosco della Città, jardim botânico próximo de Trento (BARLACCHI *ET AL.*, 2015).

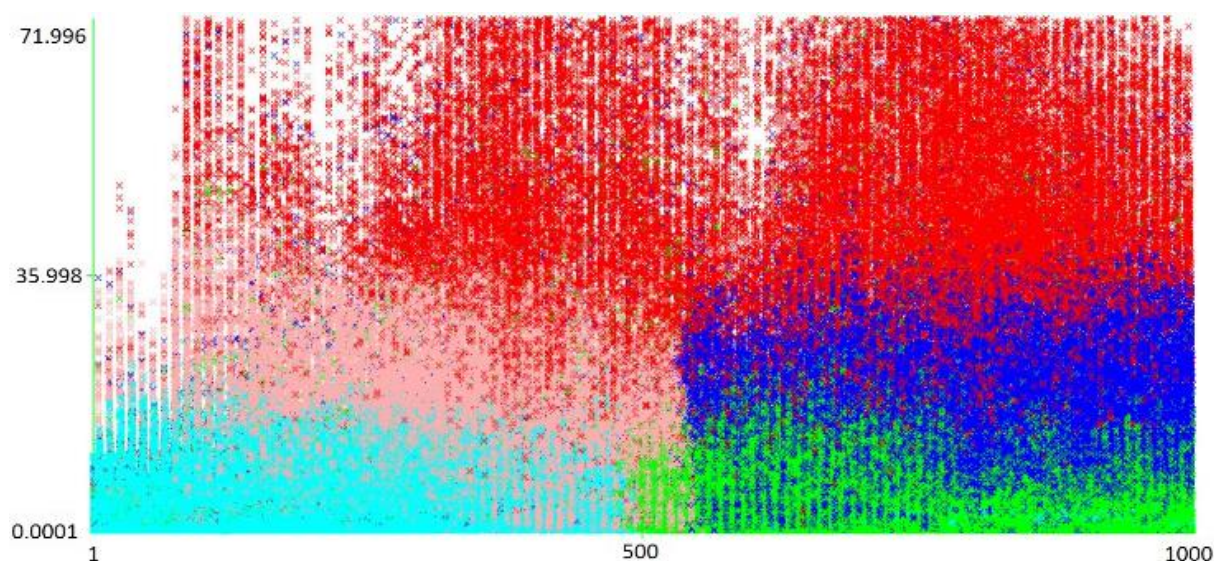


Figura 6. Total de chamadas telefônicas por antena (região)

Fonte: elaborada pelos autores

4.2 ANÁLISE TEMPORAL

Para análise temporal, verificou-se a distribuição dos tipos de serviços (chamadas telefônicas, SMS e Internet), por dias da semana, contabilizando-se o total de cada tipo de serviço em toda a rede, conforme ilustra a Figura 7, considerando-se todas as áreas geográficas da base.

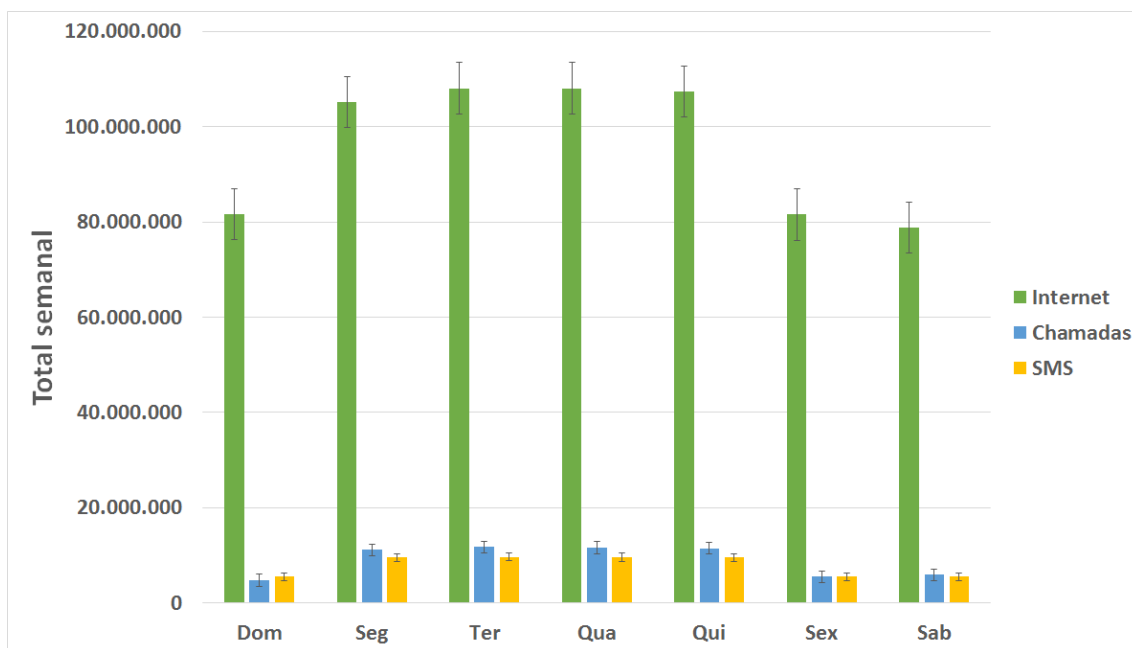


Figura 7. Distribuição dos tipos de serviços por dia

Fonte: elaborada pelos autores

Pelo que mostram os resultados, o dia com maior demanda foi a terça-feira, seguido de quarta e quinta-feira. Pela análise, verifica-se que há um forte componente sazonal, indicando rotinas mais intensas de trabalho, estudos e negócios nesses dias. Os dias mais próximos dos finais de semana, segunda e sexta, por sua vez, podem ter sido afetados justamente pela proximidade do fim e início da semana, causadas por um ritmo menor de atividades representadas pelos padrões comportamentais humanos temporais.

Após a identificação da terça-feira como dia de maior demanda, procurou-se encontrar o comportamento por faixa de horário considerando todos os tipos de serviços. A Figura 8 demonstra que o maior período de requisições de serviços compreende dois intervalos: o primeiro intervalo entre 12 e 15 horas e o segundo intervalo entre 17 e 19 horas.

Com isso, verifica-se que há um componente sazonal forte durante certos intervalos do dia, que pode ter sido influenciado com atividades rotineiras diárias. Observa-se que a demanda geralmente começa quando as pessoas acordam, geralmente às 07 horas, e aumenta durante o dia. A partir das 19 horas a demanda diminui, provavelmente, pelo fato de quando as pessoas retornam do trabalho para casa e se preparam para o descanso noturno.

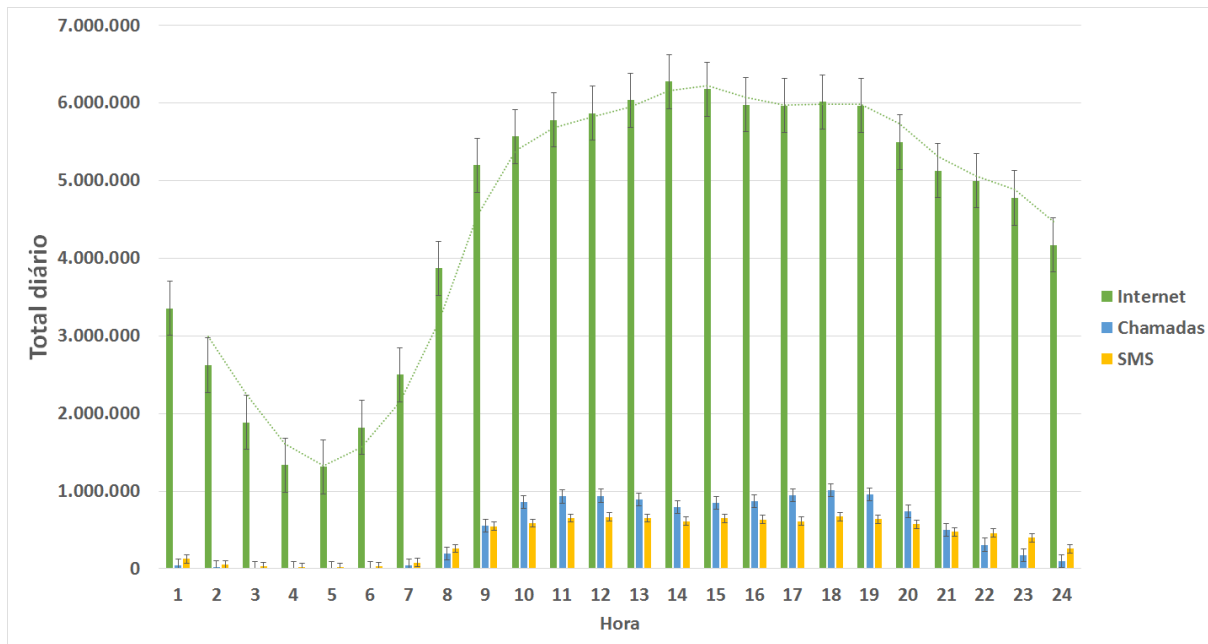


Figura 8. Distribuição dos serviços por horário durante a 3ª feira

Fonte: elaborada pelos autores

Com os resultados obtidos pela mineração de dados aplicada na rede de telefonia móvel, espera-se um melhor planejamento e a entrega de serviços móveis de qualidade nas cidades inteligentes suportadas por redes de nova geração 5G. Como os gráficos mostraram, através das análises realizadas, pode-se interferir em configurações da rede.

5 CONCLUSÃO E TRABALHOS FUTUROS

Conforme apresentado neste trabalho, há uma perspectiva da mineração de dados orientada à computação urbana em ambientes de *Big Data*. No entanto, *Big Data* apresenta enormes desafios em relação à mineração de dados, integração de dados de diversas fontes com uso de sensores e dispositivos móveis e descoberta de conhecimento. A análise de grandes volumes de dados em tempo real é um grande desafio, estimando uma análise diária da ordem de 6.8 PB de dados em uma rede móvel de uma operadora.

A principal contribuição deste trabalho foi a proposta de um *framework* baseado em mineração de dados como estrutura necessária em redes 5G. Acredita-se que, sem ferramentas como as propostas aqui, a análise de comportamento da rede será inviável devido à grande quantidade de dados que serão gerados.

Como trabalhos futuros, sugerem-se a realização de coleta de dados locais, a adoção de novas técnicas e estudos de mobilidade de usuários dentro da rede. A investigação de descobertas de padrões através da abordagem CoUrbD2M juntamente com a implantação do núcleo de rede com o *framework* CoUrbF5G proposto, permitirá que futuras redes 5G sejam capazes de se orientar pelo contexto e aprendam com a experiên-

cia. Assim, as melhores decisões de alocação de recursos e planejamento urbano poderão ser tomadas.

6 AGRADECIMENTOS

O presente trabalho é uma versão estendida e revisada do artigo “CoUrbD2M: Mineração de Dados Orientada à Computação Urbana em cenários de *Big Data* e Redes 5G”, publicado no Primeiro Workshop de Computação Urbana (CoUrb) promovido pelo Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2017).

Os autores agradecem à CAPES, ao CEFET-MG e à PUC Minas pelo apoio financeiro.

REFERÊNCIAS

AGYAPONG, P. K.; IWAMURA, M.; STAEHLE, D.; KIESS, W.; BENJEBBOUR, A. Design considerations for a 5G network architecture. *Communications Magazine (CM)* - In *Institute of Electrical and Electronics Engineers (IEEE)*, Finland, v. 52, n. 11, p. 65-75, 2014.

AYDIN, B.; ANGRYK, R. A. A graph-based approach to spatiotemporal event sequence mining. *16th Intl Conf on Data Mining Workshops (DMW)* - In *IEEE*, Barcelona, p. 1090-1097, 2016.

BARLACCHI, G.; DE NADAI, M.; LARCHER, R.; CASELLA, A.; CHITIC, CRISTIANA; TORRISI, G.; ANTONELLI, F.; VESPIGNANI, A.; PENTLAND, A.; LEPRI, B. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific Data 2*, n. 150055, p. 1-15, 2015. Disponível em: <http://www.nature.com/articles/sdata201555>. Acesso em: set/2016.

BAYIR, M. A.; EAGLE, N.; DEMIRBAS, M. Discovering spatiotemporal mobility profiles of cellphone users. 10th Intl Symp on World of Wireless, Mobile and Multimedia Networks (WoWMoM) - In *IEEE*, Buffalo and Cambridge, p.1-9, 2009.

CHEN, F.; DENG, P.; WAN, J.; ZHANG, D.; VASILAKOS, A.s V.; RONG, X. Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*, n. 431047, p. 1-14, 2015a.

CHEN, M.; ZHANG, Y.; HU, L.; TALEB, T.; SHENG, Z. Cloud-based Wireless Network: Virtualized, Reconfigurable, Smart Wireless Network to Enable 5G Technologies. *Mobile Netw Appl*, n. 20, p. 704-712, 2015b.

EAGLE, N.; PENTLAND, A.; LAZER, D. Inferring Social Network Structure using Mobile Phone Data. *Proceedings of the National Academy of Sciences*, Cambridge, v. 106, n. 36, p. 15274-15278, 2009.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM, New York*, v. 39, n. 11, p. 27-34, 1996.

GABER, M. M.; GAMA, J.; KRISHNASWAMY, S.; GOMES, J. B.; STAHL, F. Data stream mining in ubiquitous environments: state-of-the-art and current directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 4, n. 2, p. 116-138, 2014.

HE, Y.; YU, F. R.; ZHAO, N.; YIN, H.; YAO, H.; QIU, R. C. Big data analytics in mobile cellular networks. *Special Section on Theoretical Foundations for Big Data Applications: Challenges and Opportunities - In IEEE Access*, vol. 4, p. 1985-1996, 2016.

I, CHIH-LIN; LIU, Y.; HAN, S.; WANG, S.; LIU, G. On big data analytics for greener and softer RAN. *Special Section on Big Data for Green Communications and Computing - In IEEE, Beijing*, v. 3, p. 3068-3075, 2015.

ITU – International Telecommunication Union. Recommendation ITU-R M.2380-0: IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond. 2015. Disponível em: <http://www.itu.int/rec/R-REC-M.2083>. Acesso em: mar/2017.

MALIK, A.; QADIR, J.; AHMAD, B.; YAU, K. L. A.; ULLAH, U. QoS in IEEE 802.11 - based wireless networks: a contemporary review. *Journal of Network and Computer Applications*, 55, p. 24-46, 2015.

MONTEVECCHI, A. L. D. Pictorea: um método para descoberta de conhecimento em bancos de dados convencionais. Dissertação (Mestrado Acadêmico em Informática). Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2012.

MORE, J. S.; LINGAM, C. Reality mining based on social network analysis. *Intl Conf on Communication, Information & Computing Technology (CICT) - In IEEE*, Mumbai, p. 1-6, 2015.

PICTOREA. Método PICTOREA. Disponível em: <http://www.montavecchi.com.br/SPEM/>. Acesso em: set/2016.

SERRA, A. P.; ZÁRATE, L. E. Characterization of time series for analyzing of the evolution of time series clusters. In *Expert Systems with Applications*, v. 42, n. 1, p. 596-611, 2015.

WEKA, University of Waikato. Weka 3: Data Mining Software in Java. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka>. Acesso em: set/2016.

ZHENG, Y.; CAPRA, L.; WOLFSON, O.; YANG, H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, v. 5, n.3, p. 38:1-38:55, 2014.