

# CoUrbD2M: Mineração de Dados Orientada à Computação Urbana em cenários de *Big Data* e Redes 5G

Carlos Renato Storck<sup>1,2</sup>, Edwaldo Araújo Sales<sup>2</sup>,  
Luis Enrique Zárate<sup>2</sup>, Fátima de L. P. D. Figueiredo<sup>2</sup>

<sup>1</sup>Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

<sup>2</sup>Pontifícia Universidade Católica de Minas (PUC Minas)

storck@contagem.cefetmg.br, ed.araujo@outlook.com,

{zarate, fatimafig}@pucminas.br

**Abstract.** *This paper presents a data mining view as support on the urban computing in Big Data environments. Two approaches are proposed: CoUrbD2M (Urban Computing Driven Data Mining) and CoUrbF5G (Urban Computing Framework in 5G Networks) boosted by future 5G networks, fifth generation cellular networks. The objective is the demonstration of the applicability to the quality of urban life improvement, through the mobile services. The work sought to find and to analyze the real standards of a mobile telephone network, applying data mining techniques, in combination with aiding methods on the driving of Data Science processes such as KDD (Knowledge Discovery in Databases) and Big Data, in order to allow the better urban planning decisions making and the delivering of quality mobile services in smart cities supported by new generation networks.*

**Resumo.** *Este artigo apresenta uma visão da mineração de dados como apoio na computação urbana em ambientes de Big Data. São propostas duas abordagens: CoUrbD2M (Urban Computing Driven Data Mining) e CoUrbF5G (Urban Computing Framework in 5G Networks) impulsionadas pelas futuras redes 5G, redes de celulares de quinta geração. O objetivo é demonstrar aplicabilidade para melhoria da qualidade de vida urbana, através de serviços móveis. O trabalho procurou encontrar e analisar padrões reais de uma rede de telefonia móvel, aplicando técnicas de mineração de dados, em conjunto com métodos auxiliares na condução de processos de Data Science como KDD (Knowledge Discovery in Databases) e Big Data, a fim de permitir a tomada de melhores decisões de planejamento urbano e a entrega de serviços móveis de qualidade nas cidades inteligentes suportadas por redes de nova geração.*

## 1. Introdução

O planejamento urbano em cidades inteligentes requer a implantação de Redes de Nova Geração (RNG) e abordagem de técnicas e mecanismos de Computação Urbana (CoUrb) em ambientes de *Big Data*. Tais mecanismos permitem a entrega de melhores serviços para a cidade, melhorando, conseqüentemente, a qualidade de vida dos cidadãos [Zheng et al. 2014].

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. Os dados são coletados por diversas fontes espalhadas pela cidade, como sensores e atuadores, redes veiculares, dispositivos móveis, redes de telefonia, dentre outros. Nesse sentido, justifica-se o uso da mineração de dados como apoio para a descoberta de padrões preditivos não óbvios, em grandes bases de dados coletados pelas redes 5G e cenários de *Big Data*. 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 e interpretação [Fayyad et al. 1996].

Neste trabalho, foi verificado o comportamento real de uma rede de telefonia móvel, através de técnicas de mineração de dados em conjunto com um método auxiliar na condução de processos em *Data Science*, para subsidiar a CoUrb suportada pelas futuras redes 5G. Através da mineração de dados e da descoberta de padrões de condições da rede, o sistema pode fornecer entradas para 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. As ferramentas CoUrbD2M e CoUrbF5G são apresentadas. CoUrbD2M possibilita a mineração de dados orientada à computação urbana e o *framework* CoUrbF5G que é composto por módulos que, através da observação ciente de contexto e aprendizagem pela experiência, permite a análise do comportamento e das políticas e diretrizes definidas pelas prefeituras e operadoras, alocando recursos de forma dinâmica e adaptativa.

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

## 2. Principais Conceitos e Trabalhos Relacionados

Nesta seção serão apresentados os principais conceitos e trabalhos relacionados para a compreensão de como a mineração de dados orientada à CoUrb pode ser realizada em cenários de *Big Data* e redes 5G.

### 2.1. *Data Mining* e *Big Data*

A análise de séries temporais compreende métodos e técnicas de análise de dados de séries de tempo, a fim de extrair estatísticas significativas e outras características dos dados. A mineração de dados espaço-temporais é o processo de descoberta de padrões em dados espaciais associadas com o tempo. Ela 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 and Angryk 2016].

Os padrões mais comuns utilizados na mineração de dados espaço-temporais são relatados por [More and Lingam 2015]. O primeiro padrão relatado é o agrupamento espacial, considerando análise da área geográfica e podendo ser adotado o algoritmo K-means. O segundo é a análise de *outliers*. O terceiro padrão apontado é o uso de modelos

preditivos podendo adotar aprendizagem de máquina. O último padrão relatado é a busca por correlações e objetos que possuem uma alta probabilidade. Para entender o domínio do problema, são relacionadas quatro principais métricas de redes sociais: centralidade *degree*, centralidade *betweenness*, *closeness* e *eigenvalue*.

Para a mineração de padrões sequenciais móveis em uma rede de telefonia, uma proposta é apresentada por [Bradley and Rashad 2010] com o objetivo de determinar o movimento frequente padrão de um usuário móvel. Para isso, os autores utilizaram uma base real de dados contendo 100 usuários, que foi coletada em nove meses pelo MIT (*Massachusetts Institute of Technology*). Foram criados dois algoritmos denominados *Fixed-MobileSPADE* e *Dynamic-MobileSPADE*, ambos baseados no algoritmo Apriori. Após aplicados os algoritmos na base, os autores concluíram que o algoritmo *Fixed-MobileSPADE* é mais indicado para padrões mais frequentes, e o algoritmo *Dynamic-MobileSPADE* para padrões não frequentes. Outra metodologia para extração de padrões de mobilidade baseada em nuvem para computação urbana foi apresentada por [Altomare et al. 2014, Altomare et al. 2017], mas não será abordada neste trabalho.

O estado da arte e direções correntes para mineração de fluxo de dados em ambientes ubíquos são abordados por [Gaber et al. 2014]. No *survey*, os autores tratam a mineração com uso de dispositivos e sensores para suporte de aplicações críticas, compartilhamento do conhecimento por agentes para adaptação e uso de algoritmos eficientes e adaptáveis a contextos sensíveis, tais como recursos e situação. Para apoiar a mineração de dados com base nos princípios de adaptação em dispositivos móveis, os autores desenvolveram um conjunto de ferramentas denominado *Open Mobile Miner*. A mineração colaborativa de fluxo de dados ciente de contexto é uma das características mais importantes para desenvolvimento de novas aplicações.

Em [Chen et al. 2015a], é proposta uma arquitetura de mineração de *Big Data* para dados massivos, produzidos pela Internet das Coisas (IoT), que são considerados de alto valor. Eles propõem também que a mineração de dados com algoritmos adaptados para esse cenário apoiarão esse novo ambiente. Estes algoritmos de classificação, *clustering*, análise de associação, análise de séries temporais e *outliers* servem para aplicações de *e-commerce*, indústria, saúde e governo.

Em [I et al. 2015], é estudada a análise de *Big Data* para Rede de Acesso de Rádio (RAN). Os dados são coletados através da sinalização, tráfego e ambiente. A proposta do trabalho é otimizar e operar a rede 4G de forma autônoma. No trabalho, os autores apresentam o paradigma de operação da rede baseada em *Big Data*, apresentado pela Figura 1, e fluxograma da otimização da rede inteligente baseada em dados de sinalização. Entre as características do modelo do fluxograma estão a capacidade de processamento de coleta de sinalização de 1 Gbps, definição de 296 procedimentos de sinalização de representação de dados externos durante a limpeza de dados, a extração de cerca de 100 procedimentos de comunicação com definição de mais de 400 indicadores-chave de desempenho na seleção de dados e integração (validação), armazenamento de dados na ordem de 1 milhão por minuto, a verificação do modelo com 400 entradas durante a mineração de dados, e por fim 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.

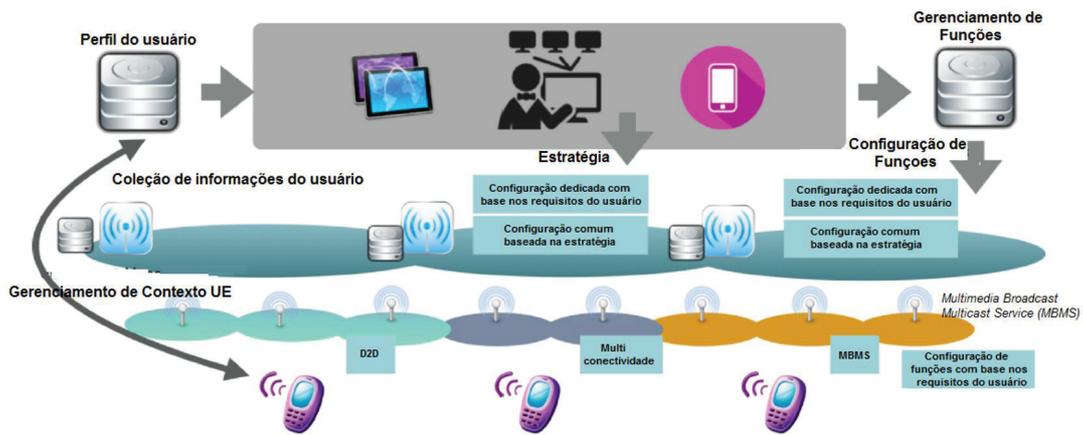


Figura 1. Paradigma e fluxograma de operação da rede baseada em *Big Data* (Adaptada de [1 et al. 2015]).

Uma arquitetura da rede sem fio na nuvem com centros de *Big Data* é proposta por [Chen et al. 2015b], apresentada pela Figura 2. 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 ficaria responsável pela cobertura, realocação e atualização de recursos, e detecção de falhas. Para QoE (*Quality of Experience*), a otimização com base em dados no comportamento ficaria responsável pela otimização de dispositivos e localização dos mesmos.

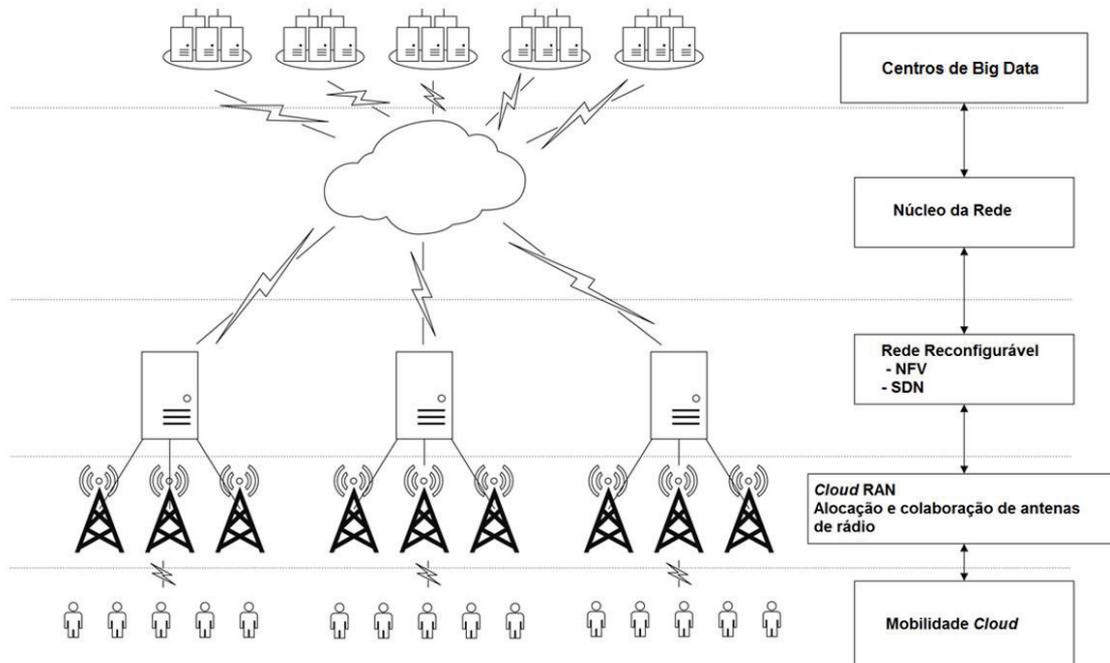
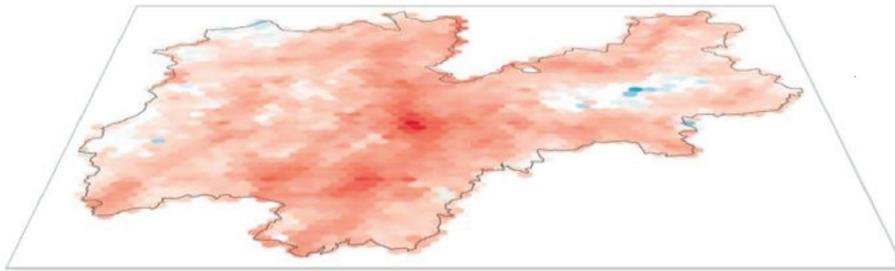


Figura 2. Arquitetura da rede sem fio na nuvem com centros de *Big Data* (Adaptada de [Chen et al. 2015b]).

Em se tratando de *Big Data*, como incentivo a comunidade acadêmica para superar os seus 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*”<sup>1</sup>. O *dataset* apresenta chamadas telefônicas, serviços SMS (*Short Message System*) e uso de Internet por horário, através de CDRs (*Call Detail Records*). Um mapa de calor dos serviços de telecomunicações da base de dados disponibilizada também é apresentado no artigo, conforme a Figura 3.



**Figura 3. Mapa de calor dos serviços de telecomunicações (Adaptada de [Barlacchi et al. 2015]).**

Embora, [Barlacchi et al. 2015] tenham verificado padrões de comportamento semanal por tipo de serviço, o que será comparado na Seção 4, no entanto, os autores não tratam a mineração de dados; que, diferentemente deste trabalho, será tratado em função do planejamento urbano para uma melhor alocação de recursos da rede.

## 2.2. Redes 5G

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

Uma característica notável das redes é a plataforma necessária das redes futuras das operadoras, a partir de 2020; que vai precisar de uma arquitetura 5G, 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], em duas camadas: uma RAN e uma rede em nuvem. Nesta 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*). Segundo os autores, diferentes tipos de estações bases e RRUs (*Remote Radio Unit*) realizam um conjunto mínimo de funções L1 / L2 da rede de rádio. A rede em nuvem, por sua vez, é constituída por uma entidade *U-plane* (UPE) e uma entidade *C-Plane* (CPE), que executam as funções das camadas mais altas, relacionadas a estas entidades.

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

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 e o controle, e possuir funções de rede virtualizadas; bem como, 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], a ciência de contexto é relatada como uma forma de melhorar tanto a qualidade de serviço, como a qualidade de experiência. 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 estes 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. CoUrbD2M e CoUrbF5G

Neste trabalho, a mineração de dados orientada à computação urbana, ou CoUrbD2M, foi desenvolvida tendo como base as diversas etapas propostas pelo método PICTOREA, descrito por [Montevecchi and Zárate 2012, Pictorea 2012]. O método serviu de base para auxílio na condução e integração da percepção humana no processo de *Data Science*. A proposta da abordagem é analisar e verificar o comportamento real de uma rede de telefonia móvel, aplicando técnicas de mineração de dados, como pré-processamento e *cluster*. Para o contexto deste trabalho, CoUrbD2M foi limitada à aplicação de métodos padrão de mineração de dados.

Sendo assim, adotou-se um conjunto de dados reais de telefonia móvel, coletado durante uma semana em Milão, e região de Trentino, durante o mês de novembro de 2013, disponibilizado pela companhia Telecom Itália no formato ODbL, para a competição *Big Data Challenge*, conforme descrito na seção anterior. A base é composta por mais de 15 milhões de registros, tendo cada CDR realizado com atribuição da antena ou Estação Rádio Base (ERB) e o tempo gasto por serviço. A base de dados foi formada pelos atributos de recebimento e realização de chamadas telefônicas, recebimento e envio de SMS, e uso de Internet. Ressalta-se que para o uso de Internet, registrou-se um CDR a cada 15 minutos ou a cada 5 MB gastos. A ferramenta Weka foi usada para analisar o conjunto de dados [Weka 2016].

Para apoiar e tornar possível a CoUrbD2M, o *framework* CoUrbF5G proposto, se

baseia em CWN para redes 5G, apresentado pela Figura 4, possuindo como entrada a abordagem de mineração de dados orientada à computação urbana. O CoUrbF5G é composto por módulos que, através da observação ciente de contexto e aprendizagem pela experiência, permite ao sistema ser capaz de analisar o comportamento e consultar políticas e diretrizes definidas pelas prefeituras e operadoras, alocando recursos de forma dinâmica e adaptativa. Além disso, o torna capaz de realizar um adequado planejamento urbano e verificar padrões ou previsões de demandas futuras das cidades; como, por exemplo, uma grande quantidade de pessoas presentes em um determinado evento, como *shows* ou jogos em estádios.

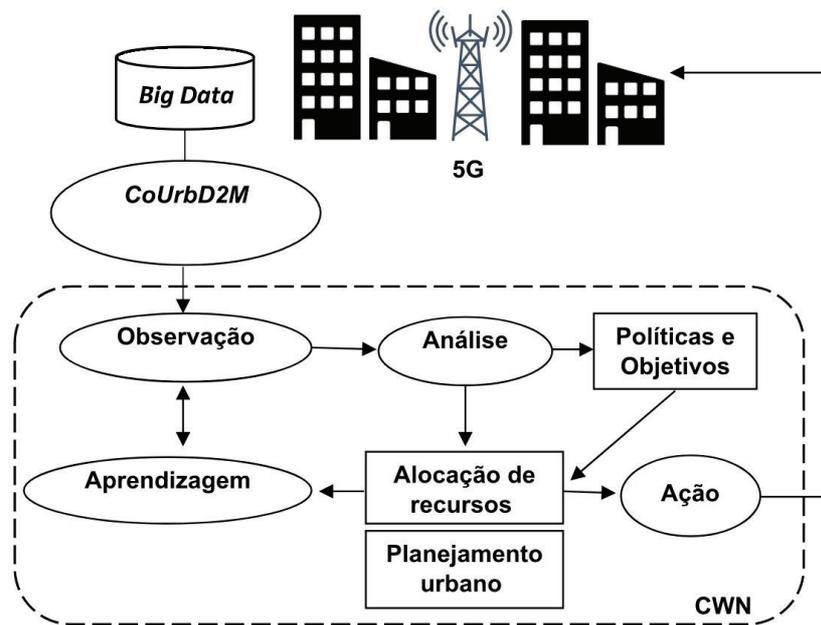


Figura 4. Representação do *framework* CoUrbF5G.

Conforme apresentado na Figura 4, o *framework* CoUrbF5G proposto 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. Ressalta-se que essas duas etapas são fundamentais, pois embasarão as etapas futuras do *framework*. 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, capacitando, então, o sistema para aprender com a experiência, juntamente com o módulo de observação.

A primeira etapa do CoUrbD2M é a exploração do espaço 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, como 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 (ERB), bem como a similaridade entre os serviços prestados, objetivando um correto planejamento futuro.

**Tabela 1. Matriz de problemas *Pairwise***

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

A próxima etapa sugerida é a definição do espaço solução. Nesta 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, o agrupamento.

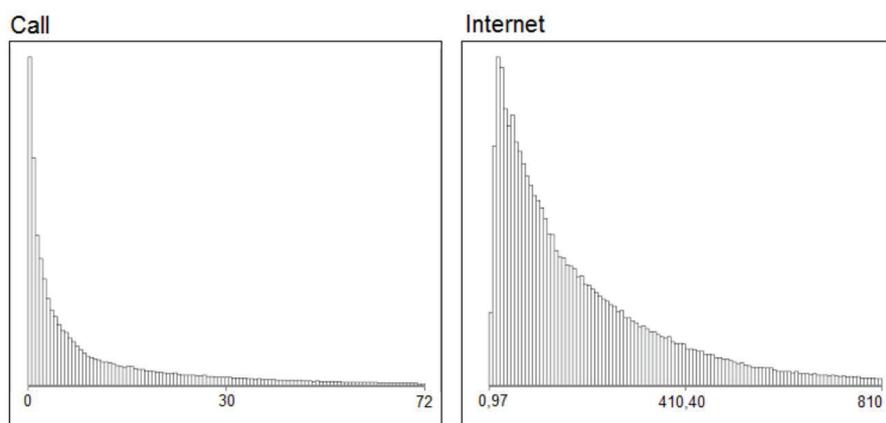
Para compreensão do domínio do problema e sua caracterização, foram verificadas quais características enfatizariam o conhecimento útil e não óbvio, podendo então ser avaliado usando os conceitos e aspectos D3M (*Domain Driven Data Mining*). A caracterização do problema através de atributos identificou 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 CSV (*Comma-separated values*) através de *scripts* na linguagem C#, verificando sua dimensionalidade. Foram verificadas: a consistência e coerência dos atributos, das instâncias e 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. Procurou-se obter a média, desvio padrão, distribuição probabilística e análise de correlação, para a construção dos histogramas dos atributos de eventos realizados ao longo do tempo apresentados pela Figura 5.

Na etapa de preparação dos atributos, foi verificado os valores ausentes e análise de *outliers* presentes na base de dados. Para valores ausentes, optou-se por eliminar todos os registros encontrados. Para análise de *outliers*, aplicou-se *cluster*, conforme os resultados apresentados na Seção 4

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 considerar o tempo de serviço independente de envio ou recebimento. O mesmo ocorreu com os atributos *call-in* e *call-out*. Neste trabalho, foi utilizada uma amostra representativa de instâncias do banco de dados, que resultou em um primeiro subconjunto de dados formado por 499.999 instâncias, sendo que a base original continha 15.085.579 registros CDRs. Para fins de comparação na análise de *outliers*, utilizou-se também um novo subconjunto contendo 2.008.105 instâncias.



**Figura 5. Histogramas dos atributos *call* e *internet*.**

Na etapa de pré-processamento, algoritmos de mineração de dados foram executados previamente. Como todos os dados da base são numéricos, foi verificado que a mudança de escala poderia ser aplicada, garantindo a preservação das características dos valores originais.

Em seguida, foram executadas as técnicas de mineração de dados, cujos padrões foram documentados para verificação qualitativa do conhecimento descoberto. Os resultados estão apresentados na próxima seção, permitindo que o especialista de domínio da computação urbana valide os dados e verifique se atendeu as expectativas esperadas, conforme propõe o método de base PICTOREA.

#### 4. Resultados

Os resultados da análise obtida nos experimentos são apresentados nesta Seção.

Conforme descrito na Seção 3, o *dataset* foi explorado e realizou-se o pré-processamento através da ferramenta Weka. Para o agrupamento de dados, foi usado primeiramente o algoritmo K-means com dois (Tabela 2) e cinco *clusters* (Tabela 3), para fins de comparação. A adoção de cinco clusters como medida de avaliação do agrupamento foi empregada de acordo com as indicações de More e Lingam [More and Lingam 2015].

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

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

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

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

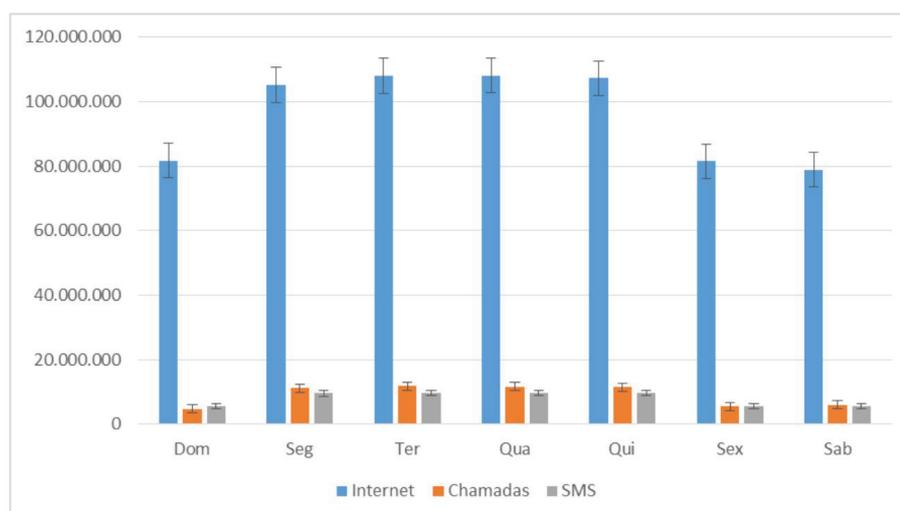
No segundo momento, o algoritmo EM (*Expectation Maximization*) que é um método baseado em modelos foi executado para fins de comparação com K-means, conforme apresentado na Tabela 4.

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

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

Após executado o agrupamento, aplicou-se a média com dois desvios padrão na análise de *outliers*, sendo encontrados 60.313 registros CDRs, considerados fora do padrão; o que representa 12,06% do primeiro subconjunto de dados. Em um segundo experimento, verificando um subconjunto de dados de 2.008.105 instâncias, foram encontradas 97.840 instâncias, consideradas *outliers*; o que representa 4,87% deste segundo subconjunto de dados.

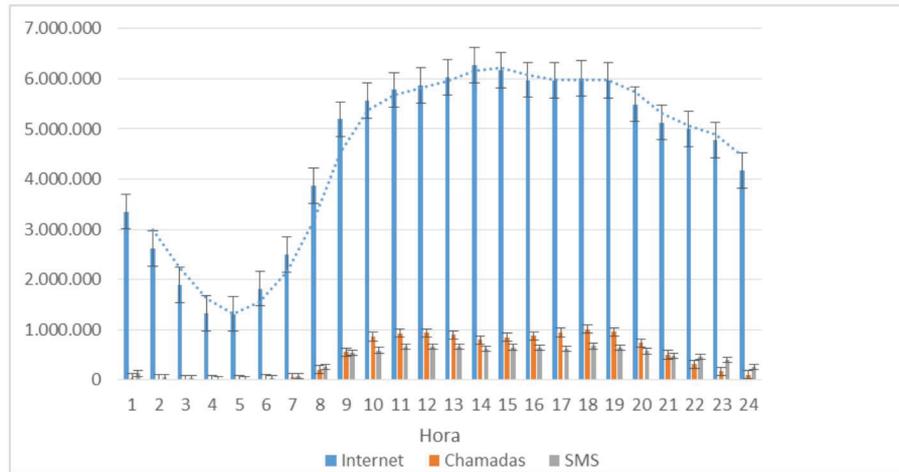
Para verificar a abordagem no primeiro subconjunto de dados e comparação da análise temporal, como apresentado por [Barlacchi et al. 2015], que considerou a contagem de eventos na Cidade de Milão, constatando que a maior demanda seria na sexta-feira; verificou-se a distribuição dos tipos de serviços (chamadas telefônicas, SMS e Internet), por dias da semana, contabilizando o tempo total gasto de cada evento em toda a rede e não somente o número de conexões, conforme a Figura 6, considerando todas as áreas geográficas da base e o tempo gasto por tipo de serviço. O dia com maior demanda foi a terça-feira, seguidos de quarta-feira e quinta-feira, respectivamente.



**Figura 6. Distribuição dos tipos de serviços por dia.**

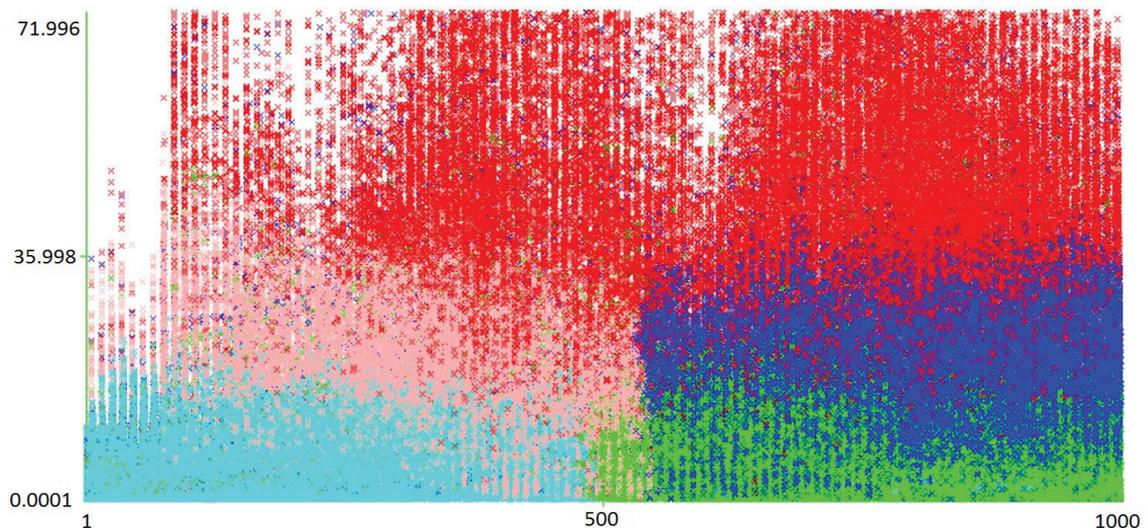
Após a identificação da terça-feira como dia de maior demanda, procurou-se encontrar o comportamento por faixa de horário considerando todos os tipos de serviços. A Figura 7 demonstra que o maior período de requisições de serviços compreende em dois

intervalos: o primeiro intervalo entre 12 e 15 horas, e o segundo intervalo entre 17 e 19 horas.



**Figura 7. Distribuição dos serviços por horário durante a 3ª feira.**

Na análise espacial, em [Barlacchi et al. 2015] foi apresentado o mapa de calor, porém não foi apresentada a demanda por tipo de serviço. Sendo assim, a Figura 8 representa a demanda de cada antena (ERB) para chamadas telefônicas agrupadas em cinco *clusters* de acordo com sua similaridade. Da mesma forma, podem ser visualizadas e observadas as demandas por outros serviços, tais como SMS e Internet.



**Figura 8. Demanda de chamadas telefônicas por antena.**

Com os resultados obtidos pela mineração de dados aplicada na rede de telefonia móvel, entregues e tratados pela implementação do *framework* CoUrbF5G, espera-se um melhor planejamento urbano e a entrega de serviços móveis de qualidade nas cidades inteligentes suportadas por redes de nova geração 5G.

## 5. Conclusões e Trabalhos Futuros

Conforme apresentado neste trabalho, investigou-se a adoção de técnicas para um planejamento urbano inteligente através de KDD e mineração de dados 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 do grande volume 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.

As principais contribuições deste trabalho foram as propostas de mineração de dados orientada à computação urbana e sua estrutura necessária de computação urbana em redes 5G, através das ferramentas CoUrbD2M e CoUrbF5G. Foram mostrados como os dados obtidos, através de mineração, podem auxiliar no planejamento das redes de nova geração 5G.

Como trabalhos futuros, sugere-se a realização de coleta de dados locais, 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* proposto, permitirá que futuras redes 5G sejam capazes de se orientar pelo contexto, aprendam com a experiência e tomem as melhores decisões de alocação de recursos e planejamento urbano.

### Agradecimentos

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., and Benjebbour, A. (2014). Design considerations for a 5g network architecture. *IEEE Communications Magazine*, 52(11):65–75.
- Altomare, A., Cesario, E., Comito, C., Marozzo, F., and Talia, D. (2014). Trajectory pattern mining over a cloud-based framework for urban computing. In *High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICSS), 2014 IEEE Intl Conf on*, pages 367–374. IEEE.
- Altomare, A., Cesario, E., Comito, C., Marozzo, F., and Talia, D. (2017). Trajectory pattern mining for urban computing in the cloud. *IEEE Transactions on Parallel and Distributed Systems*, 28(2):586–599.
- Aydin, B. and Angryk, R. A. (2016). A graph-based approach to spatiotemporal event sequence mining. In *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*, pages 1090–1097. IEEE.
- Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., Antonelli, F., Vespignani, A., Pentland, A., and Lepri, B. (2015). A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data*, 2.
- Bayir, M. A., Demirbas, M., and Eagle, N. (2009). Discovering spatiotemporal mobility profiles of cellphone users. In *World of Wireless, Mobile and Multimedia Networks &*

- Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, pages 1–9. IEEE.
- Bradley, J. G. and Rashad, S. S. (2010). Mining mobile sequential patterns in wireless cellular networks. In *Technological Developments in Networking, Education and Automation*, pages 597–602. Springer.
- Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., and Rong, X. (2015a). Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*.
- Chen, M., Zhang, Y., Hu, L., Taleb, T., and Sheng, Z. (2015b). Cloud-based wireless network: Virtualized, reconfigurable, smart wireless network to enable 5g technologies. *Mobile Networks and Applications*, 20(6):704–712.
- Eagle, N., Pentland, A. S., and Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34.
- Gaber, M. M., Gama, J., Krishnaswamy, S., Gomes, J. B., and Stahl, F. (2014). Data stream mining in ubiquitous environments: state-of-the-art and current directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(2):116–138.
- I, C. L., Liu, Y., Han, S., Wang, S., and Liu, G. (2015). On big data analytics for greener and softer ran. *IEEE Access*, 3:3068–3075.
- ITU (2015). Recommendation itu-r m.2380-0: Imt vision – framework and overall objectives of the future development of imt for 2020 and beyond. <http://www.itu.int/rec/R-REC-M.2083>.
- Malik, A., Qadir, J., Ahmad, B., Yau, K.-L. A., and Ullah, U. (2015). Qos in ieee 802.11-based wireless networks: a contemporary review. *Journal of Network and Computer Applications*, 55:24–46.
- Montevecchi, A. L. D. and Zárata, L. E. (2012). 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.
- More, J. S. and Lingam, C. (2015). Reality mining based on social network analysis. In *Communication, Information & Computing Technology (ICCICT), 2015 International Conference on*, pages 1–6. IEEE.
- Pictorea (2012). Método pictorea. <http://www.montevecchi.com.br/SPEM/>.
- Weka (2016). Data mining - weka. <http://community.pentaho.com/projects/data-mining/>.
- Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38.