



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Presidente Prudente

ANÁLISE E VISUALIZAÇÃO DE DADOS DA FIFA WORLD CUP 2014
ORGANIZADOS EM GRAFOS E DIGRAFOS

TEORIA DOS GRAFOS
PROFº. DRº. DANILO MEDEIROS ELER

BRUNO SANTOS DE LIMA
LEANDRO UNGARI CAYRES

PRESIDENTE PRUDENTE
AGOSTO - 2016

Sumário

1. Introdução	3
2. Conjunto de dados.....	3
3. Organização do conjunto de dados	4
4. Exploração do conjunto de dados	7
5. Referências Bibliográficas.....	12

1. Introdução

Esta coletânea relata a realização de análise e exploração sobre um conjunto de dados, onde este último está presente no mundo real e consequentemente faz parte do dia-a-dia de alguma parcela da população mundial, este conjunto foi modelado e organizado para uma estrutura afim de representar um grafo, ou mesmo um dígrafo, posteriormente foi aplicado algoritmos da frente de teoria dos grafos com o objetivo de explorar esses dados afim de obter algumas informações, muitas vezes de caráter estatístico.

Este relatório encontra-se organizado como segue: na Seção 2 é apresentado o conjunto de dados que foi escolhido pelo grupo, na Seção 3 é detalhado como este conjunto de dados foi modelado na forma de grafos e dígrafos, por fim na Seção 4 é discutido o processo de exploração desse conjunto de dados.

2. Conjunto de dados

O conjunto de dados escolhido para a realização da análise e exploração do mesmo foi a FIFA World Cup edição de 2014 (Copa do Mundo FIFA 2014). A competição reúne as 32 melhores seleções de futebol do mundo para a realização de um torneio que tem como objetivo unir as nações através do esporte em um país sede e mostrar ao mundo quem é a melhor seleção nacional de futebol do mundo.

O torneio é composto por 32 seleções nacionais de futebol divididas em 8 grupos de 4 times. Na fase de grupos, todos os integrantes de um mesmo grupo jogam contra si, em partida única, os dois melhores avançam para o mata-mata. Em seguida, na fase eliminatória, os times jogam partida única promovendo o afunilamento da competição até a definição do campeão.

A competição é realizada em um único país sede, no caso da edição de 2014 o evento aconteceu no Brasil que contava com 12 cidades sedes, sendo elas: Porto Alegre, Curitiba, São Paulo, Rio de Janeiro, Belo Horizonte, Cuiabá, Brasília, Salvador, Recife, Natal, Fortaleza, Manaus.

A seleções participantes da edição de 2014 foram: Brasil, Argentina, Colômbia, Chile, Equador, Uruguai, Irã, Coreia do Sul, Japão, Austrália, Costa do Marfim, Nigéria, Camarões, Gana, Argélia, Honduras, Costa Rica, Estados Unidos, México, Bélgica, Holanda, Suíça, Alemanha, Itália, Grécia, Inglaterra, Rússia, Bósnia e Herzegovina, Croácia, Espanha, Portugal e França.

3. Organização do conjunto de dados

Com base no conjunto de dados descrito anteriormente tem-se a necessidade de modelar esses dados afim de poder representa-los através de grafos ou dígrafos para que assim possa ser realizada a exploração sobre eles.

Tendo um conjunto de dados muito amplo se tratando de uma competição como a Copa do Mundo (2014), apenas alguns deles foram escolhidos para serem explorados, sendo estes: as seleções participantes, a tabela de grupos em que cada seleção participou, quantidade de gols marcados em cada jogo, quantidade de público pagante em cada partida, qual seleção venceu determinada partida, quais foram as cidades sedes da competição, qual a distância terrestre (rodovia) entre as cidades sedes, em quais sedes cada seleção jogou.

Os dados descritos acima foram abstraídos do site oficial da Federação Internacional de Futebol – FIFA, mais especificamente das sumulas dos jogos da Copa do Mundo de 2014.

Todos esses dados foram agrupados em um conjunto de arquivos texto, a estrutura desses arquivos seguem um padrão, sendo que os dados dentro do arquivo são representados de forma numérica, assim o primeiro número informa se é um grafo ou um dígrafo, o segundo informa a quantidade de seleções ou cidades sedes essa quantidade representa o número de vértices do grafo ou dígrafo, as próximas linhas indicam três números sendo eles o vértice de origem, o vértice de destino e o peso da aresta entre os dois vértices. Observe a imagem abaixo que mostra dados referentes a fase de grupos mais especificamente o grupo A, grupo este no qual a seleção brasileira de futebol está presente:

```

1
4
0 2 2
0 3 0
0 1 3
2 1 4
3 1 1
3 0 0
3 2 2

```

Figura 1- Fase de grupos - Jogos do grupo A

Com relação aos arquivos textos que contém dados sobre os jogos da competição, mais especificamente sobre quem ganhou a partida, os arquivos com esses são representações de dígrafos, assim se observarmos na imagem anterior que mostra os jogos do grupo A, tem-se o Brasil com o código zero e a Croácia com código dois, assim temos um vértice origem no Brasil e destino na Croácia, ou seja, a direção da aresta entre esses dois vértices informa sempre o vencedor, sendo o vencedor o vértice origem e o perdedor o vértice destino, o peso da aresta informa os saldo de gols da partida, ou seja, se o peso for zero significa que ocorreu um empate assim não ouve um vencedor, se o peso for um número maior que zero significa que o vértice origem ganhou a partida com saldo de gols o valor deste peso.

Além deste tipo de arquivo temos um semelhante ao descrito acima com apenas uma modificação com relação ao peso, neste caso o peso agora informa se a seleção origem venceu a partida, se ela venceu o peso é um se a seleção empatou ou perdeu então o peso é zero. Observe a imagem abaixo:

```

1
32
0 2 1
2 0 0
3 1 1
1 3 0
0 3 0
3 0 0
1 2 1
2 1 0
0 1 1
1 0 0
3 2 1
2 3 0
7 6 1
6 7 0

```

Figura 2 – Arquivo da competição completa

O arquivo da imagem acima contém dados dos vencedores de todos os jogos da competição, incluindo a fase de grupos e a fase eliminar, os jogos são sempre apresentados em pares, no destaque em azul temos o jogo Brasil contra Croácia, no caso o Brasil sendo origem ganhou da Croácia por isso o peso um, já a Croácia como origem perdeu do Brasil assim o peso fica zero. No destaque em vermelho no jogo Brasil contra México ocorreu um empate assim ambas as linhas contém o peso zero.

```

0
32
0 2 62
3 1 39
0 3 60
1 2 39
1 0 69
2 3 41
6 7 48
4 5 40
4 7 42
5 6 74
4 6 39
5 7 62

```

Figura 3 - Público dos jogos

Acima temos uma imagem que mostra agora não mais um dígrafo, mas sim um grafo onde o peso das arestas representa a quantidade de público de uma partida e milhares, logo no jogo entre Brasil (código zero) e Croácia (código dois) teve um público de 62 mil pessoas, a imagem é apenas parte do arquivo que contém todos os jogos da fase de grupos e seus respectivos públicos.

Além dos arquivos mencionados anteriormente ainda temos outros dois arquivos textos, o primeiro deles representa um grafo que contém as doze sedes que são representadas por vértices bem como o peso das arestas, entre uma sede (vértice) e outra, representa a distância entre as duas com base em rotas rodoviárias, o outro arquivo que representa um dígrafo contém também as doze sedes representadas como vértices porem o peso de suas arestas são o código das trinta de duas seleções, assim podemos saber por onde cada seleção passou.

Cada uma das cidades sedes possui um código único, de zero a onze, que identifica cada uma das sedes de forma única, além disso cada uma das seleções sempre que forem tratadas coletivamente, ou seja, sempre que forem analisadas as 32 seleções juntas, cada uma tem um código que varia de zero a trinta e um. Se um dos grupos da fase de grupo estiver sendo analisado de forma isolada então as 4 seleções deste grupo são ordenadas por ordem alfabética e cada uma recebe um código que varia de zero a três de acordo com a ordem alfabética, este último critério só é utilizado quando analisamos um grupo isoladamente, caso contrário o código das seleções permanecem os definidos inicialmente de zero a trinta e um.

4. Exploração do conjunto de dados

O conjunto de dados foi modelado para ser representado através de grafos e dígrafos, assim é dado início a exploração deste conjunto de dados, para isso foi utilizada uma ferramenta mostrada abaixo com algumas funcionalidades. O foco desta ferramenta está no contexto de visualizar o conjunto de dados destacando visualmente alguns pontos importantes do mesmo, além de destacar visualmente algumas conclusões sobre o mesmo, outras informações e estatísticas são descritas na tela auxiliar.

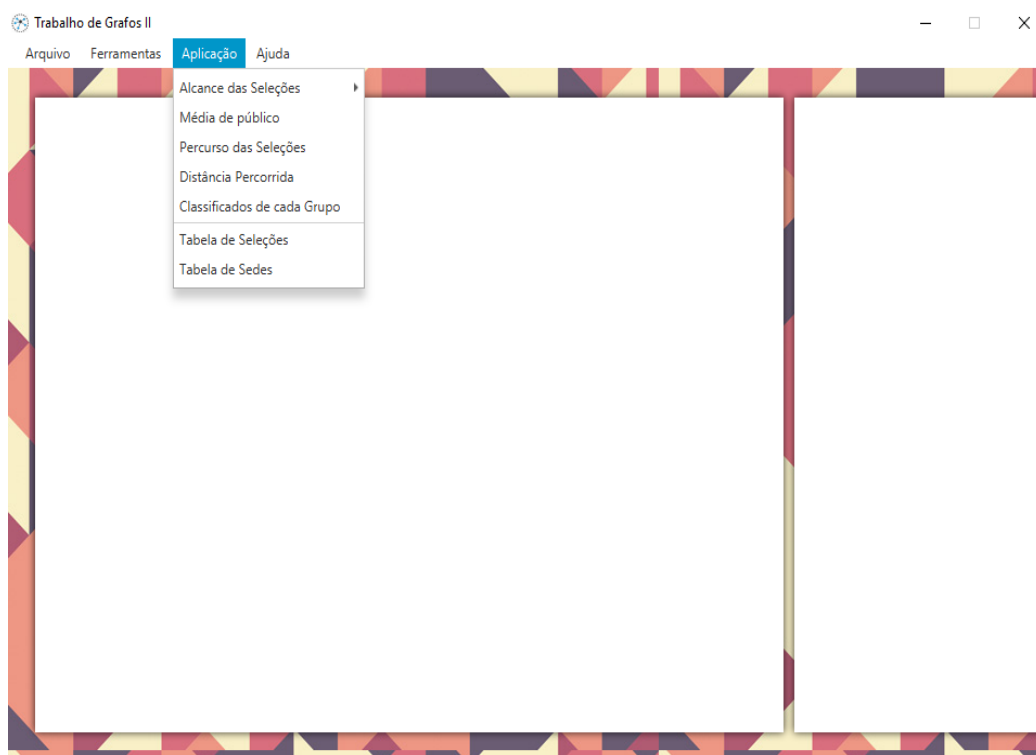


Figura 4 - Instantâneo da interface da Ferramenta com as funcionalidade para exploração dos dados

O objetivo da exploração é destacar algumas informações e responder algumas perguntas sobre o torneio, para isso alguma das questões a serem respondidas e destacadas são: até qual fase da competição cada seleção chegou? Quais seleções passaram da fase de grupos? Das oitavas de final? Das quartas de final? Da semifinal? E quais foram as duas seleções que chegaram a final mostrando a Campeã e a Vice-campeã. Mostrar a média de público da competição por completa, por grupo e por cada seleção de forma individual, além de informa quais seleções tiveram maior média de público. Outro ponto a ser destacado é qual a distância que uma determinada seleção percorreu entre as cidades sedes dos jogos durante a competição, bem como a ordem do percurso em que determinada seleção percorreu durante a competição. Observe a imagens abaixo:

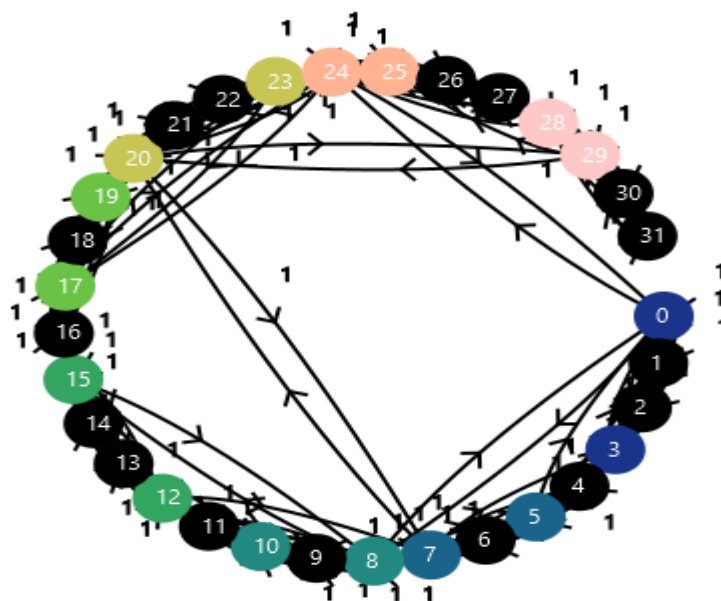


Figura 5 - Classificados para as oitavas de final

Na imagem acima os vértices com cores iguais são os classificados do grupo que avançam para a fase de oitavas de final.

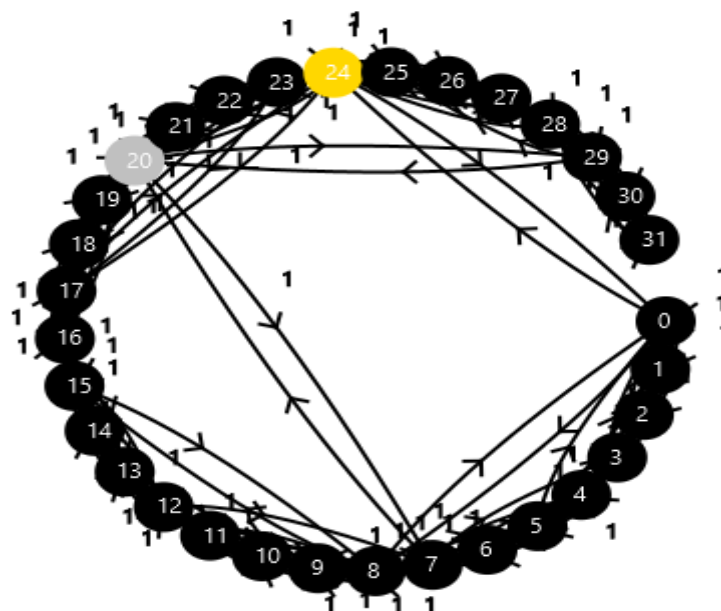


Figura 6- Seleções finalistas do Copa do Mundo 2014

Na imagem acima temos as seleções que alcançaram a final da competição, sendo o vértice na cor amarela que no caso foi a Alemanha, código 24, e a vice-campeã com a cor prata que no caso foi a Argentina, código 20.

Na figura abaixo temos o instantâneo da ferramenta analisando a média de público referente ao grupo A, logo algumas informações foram abstraídas do grafo como a média de público deste grupo, bem como a média de público de cada seleção deste grupo mostrando que no caso o Brasil, código zero, teve o maior público do grupo, seguido por Camarões, código um, Croácia, código dois e México, código três. Além disso ainda tem-se como funcionalidade a exibição de um grafo geral a abstrair a média geral de público da competição na fase de grupos, bem como as seleções com maiores públicos nesta fase.

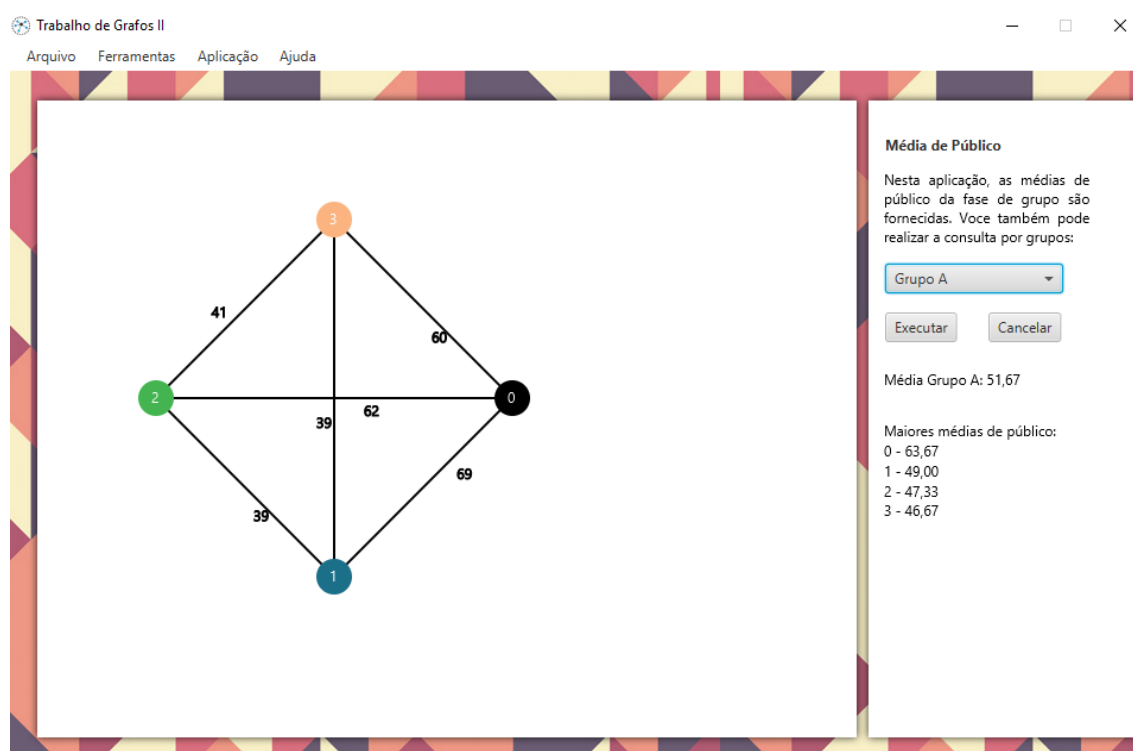


Figura 7- Instantâneo média de público do grupo A

Outro conjunto de informações que podem ser retiradas do conjunto de dados são a distância percorrida por uma determinada seleção durante a competição, bem como identificar quais sedes essa seleção passou. Observe o instantâneo abaixo que informa a distância percorrida pela seleção brasileira que foi de 11241 Km e as sedes no qual a seleção do Brasil jogou que foram: São Paulo, Fortaleza, Brasília e Belo Horizonte.

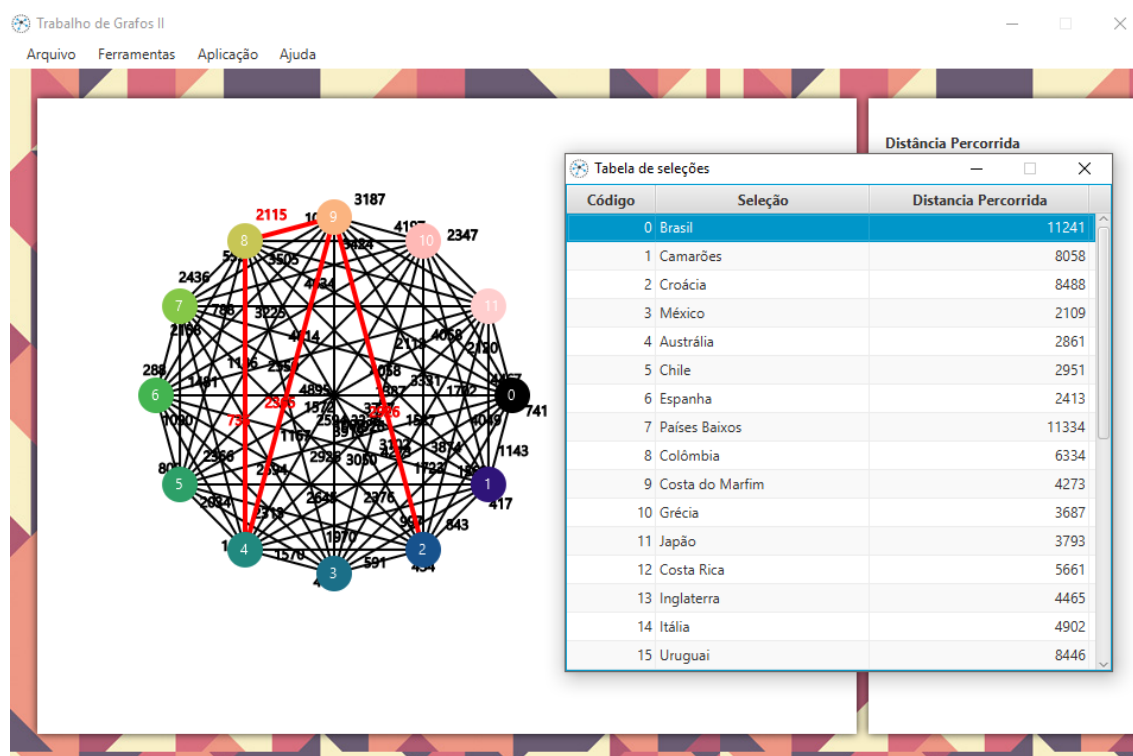


Figura 8- Instantâneo distância percorrida pela seleção brasileira

Além disso podemos também acionar outra funcionalidade para analisar qual foi a ordem do percurso em que uma seleção percorreu, ou seja, a ordem de visita das cidades sedes em que ela jogou. Observe o instantâneo abaixo analisando a seleção da Grécia quanto ao percurso: Primeiro jogo em Belo Horizonte, depois em Natal, posteriormente em Fortaleza e encerrando sua participação em Recife.

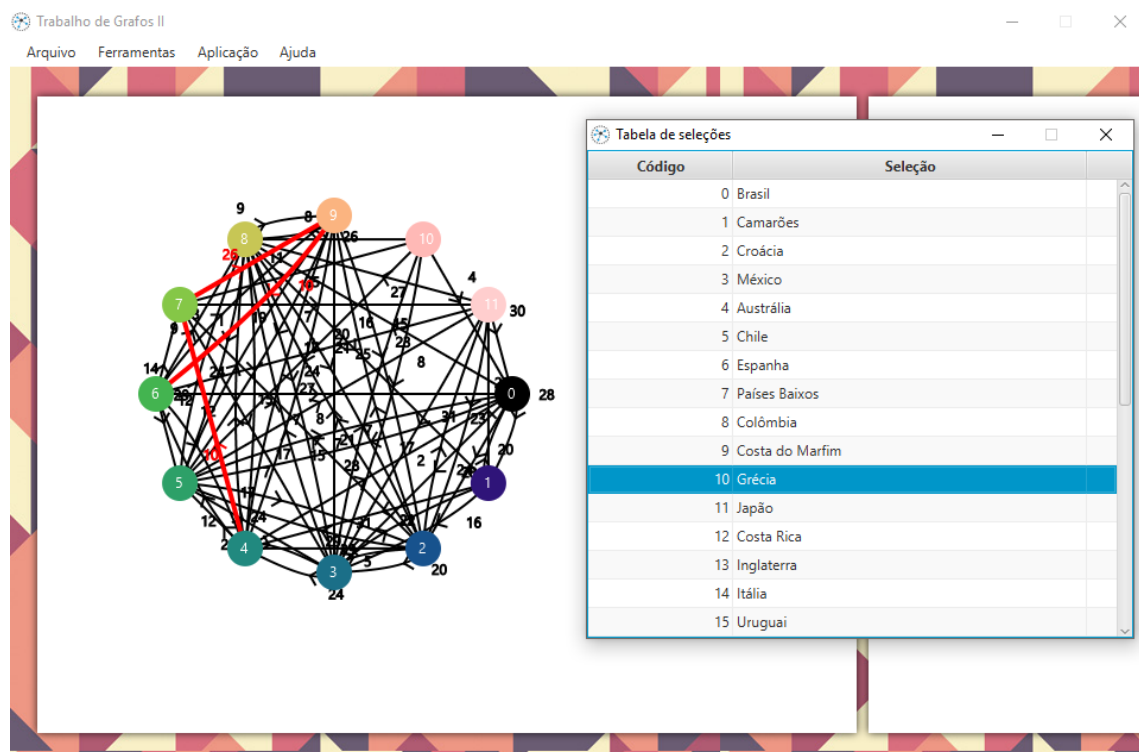


Figura 9 - Percurso da seleção grega pelas cidades sedes

5. Referências Bibliográficas

[1] FIFA, Federação Internacional de Futebol. Disponível em: < www.fifa.com.> Acesso em: 18 Ago. 2016.

[2] FIFA, Federação Internacional de Futebol. Disponível em: < www.fifa.com/worldcup > Acesso em: 18 Ago. 2016.